

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**



# WPI

(TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSPRCH.PP protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:59:27 1998; MasPar time 2.56 Seconds

Tabular output not generated.

Title: >US-08-452-843-30  
Description: (1-9) from US08452843.pep  
Perfect Score: 19  
Sequence: 1 PXXXXXXX 9

Scoring table:  
PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

a-geneseq2  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 8.237; Variance 14.300; scale 0.576

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	19	100.0	16	10	R51409	Branched peptide C5A.	1.44e+03
2	19	100.0	24	4	R22108	Peptide 5-24 derived	1.44e+03
3	19	100.0	39	10	R51402	Branched peptide 3K7	1.44e+03
4	19	100.0	60	8	R39794	Human Kunitz-type pro	1.44e+03
5	19	100.0	60	8	R39792	Human Kunitz-type pro	1.44e+03
6	19	100.0	67	7	R33871	Polypeptide p1684 com	1.44e+03
7	19	100.0	110	6	R32028	Variant IGE - mutant	1.44e+03
8	19	100.0	115	1	P90149	Sequence of hepatitis	1.44e+03
9	19	100.0	121	14	R65162	Humanised anti-PRN3	1.44e+03
10	19	100.0	123	16	R92515	Human TIRAP1.	1.44e+03
11	19	100.0	158	13	R72829	VR-2332 ORF 6 product	1.44e+03
12	19	100.0	172	10	R51467	AK beta subunit T30A.	1.44e+03
13	19	100.0	199	10	R54659	Human CD69.	1.44e+03
14	19	100.0	242	8	R41262	VWF fragment Th492-V	1.44e+03
15	19	100.0	293	14	R85251	Pseudomonas mendocina	1.44e+03
16	19	100.0	305	4	R22027	Peroxisome forming fa	1.44e+03
17	19	100.0	315	4	R20083	Maize RIP deriv. RDT-	1.44e+03
18	19	100.0	352	6	R32708	Human tau-protein.	1.44e+03

19	19	100.0	380	16	R77377	Bacillus strain DSM 5	1.44e+03
20	19	100.0	387	16	R90920	IT4 Y124P/IgG1 protei	1.44e+03
21	19	100.0	397	4	R20088	PN-I alpha analogue,	1.44e+03
22	19	100.0	398	4	R21995	PN-I beta analogue, P	1.44e+03
23	19	100.0	402	4	R20084	Maize RIP deriv. RDT-	1.44e+03
24	19	100.0	411	8	R40846	Aspartokinase II.	1.44e+03
25	19	100.0	415	6	R32922	AMP-1.	1.44e+03
26	19	100.0	464	8	R42904	Human antithrombin II	1.44e+03
27	19	100.0	464	8	R42903	Human antithrombin II	1.44e+03
28	19	100.0	467	8	R39470	mXR-alpha.	1.44e+03
29	19	100.0	474	10	R54201	Feline T cell protein	1.44e+03
30	19	100.0	485	16	R81835	Bacillus sp. alkaline	1.44e+03
31	19	100.0	501	13	R69600	New TGF-beta family m	1.44e+03
32	19	100.0	533	8	R39468	hXR-beta1.	1.44e+03
33	19	100.0	555	1	P94857	Expression plasmid pu	1.44e+03
34	19	100.0	565	14	R76699	ShET2 enterotoxin enc	1.44e+03
35	19	100.0	582	1	R76700	ELI2 enterotoxin enco	1.44e+03
36	19	100.0	582	1	P93665	Hemagglutinin-neuram	1.44e+03
37	19	100.0	585	10	R59517	Rat GAD65.	1.44e+03
38	19	100.0	585	10	R59516	Human GAD65.	1.44e+03
39	19	100.0	667	3	R11973	Tobacco SURA-Hra muta	1.44e+03
40	19	100.0	667	3	R11974	Tobacco SURA-C3 muta	1.44e+03
41	19	100.0	676	3	R15694	Thermotable T. aquat	1.44e+03
42	19	100.0	919	1	P90996	Human androgen recept	1.44e+03
43	19	100.0	1016	2	R08054	HIV-1 pol protein of	1.44e+03
44	19	100.0	1016	2	R08054	ACNPV-HIVHspol protei	1.44e+03
45	19	100.0	1146	3	R15156	Adelson Related Gene,	1.44e+03

## ALIGNMENTS

RESULT 1  
ID R51409 standard; peptide: 16 AA.  
AC R51409;  
DT 26-OCT-1994 (first entry)  
DE Branched peptide C5A.  
KW Branch: lysine dimer; lysine octamer; immunoassay; reagent;  
KW Non-A, Non-B hepatitis; NANBH; NANBH-related antibody;  
KW hepatitis C; HCV; ELISA; passive haemagglutination; vaccine.  
OS Synthetic.  
FH  
FS Key  
FT modified\_site 16  
FT location/Qualifiers  
FT /note="Linked to a lysine dimer to form branched molecule"

PD W09406826-A.  
PD 31-MAR-1994.  
PF 15-SEP-1993; U08638.  
PR 15-SEP-1992; US-946054.  
PA (UNBI-) UNITED BIOMEDICAL INC.  
PI Hosen B, Wang CY;  
PI WPI: 94-118396/14.  
DR New branched peptides for detecting non-A, non-B hepatitis antibodies  
PT - to identify carriers and diagnose hepatitis C infection, contain at  
PT least one specific epitope  
PS Claim 11, Page 12; 43pp; English.  
CC The sequences given in R51389-426 are peptides which form branched  
CC peptides through either a lysine dimer or a lysine octamer at the C-  
CC terminal. These branched peptides are immunoassay reagents for  
CC detecting Non-A, Non-B hepatitis (NANBH)-related antibodies. They  
CC may be used in the diagnosis of hepatitis C (HCV) and for the  
CC identification of carriers, particularly by ELISA or passive  
CC haemagglutination. They can also be used as vaccines to protect  
CC against HCV infection.  
SQ Sequence 16 AA;

Query Match 100.0%; Score 19; DB 10; Length 16;  
Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
DB 7 pydgmm 14  
QY 2 PXXXXXX 9

RESULT 2  
 ID R22108 standard; protein: 24 AA.  
 AC R22108; (first entry)  
 DT 15-JUL-1992  
 DE Peptide 5-24 derived from CD18 leukocyte integrin.  
 KW Cell adhesion; intercellular adhesion molecule; endothelium;  
 inflammation; leukocyte chemotaxis; rhinovirus; common cold.  
 PN W09203473-A.  
 PD 05-MAR-1992.  
 PE 23-AUG-1991; U06063.  
 PR 27-AUG-1990; US-573624.  
 PP (CETUS CORP.  
 PI Liu DY, Kaymakcalan Z, Mundy K;  
 DR WPI: 92-096833/12.  
 PT Peptide(s) derived from beta sub-unit CD18 of leukocyte integrins  
 PT - prevent leukocyte binding to ICAM and leukocyte chemotaxis, for  
 PT treating inflammatory diseases and rhinoviral infection  
 PS Claim 2; Page 23; 31pp; English.  
 CC This peptide (derived from CD18 - see R24256) was synthesised and  
 CC tested for capacity to interfere with or block adhesion of  
 CC polymorphonuclear leukocytes to human endothelial cell monolayers.  
 CC The peptide shows little inhibitory activity, c.f. peptides 4-29,  
 CC 1-26, 2-24 and 3-29. See Q22780 and R22104-R22112.  
 SQ Sequence 24 AA.

Query Match 100.0%; Score 19; DB 4; Length 24;  
 Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 13 pegidam 20  
 QY 2 PXXXXXXM 9

RESULT 3  
 ID R51402 standard; peptide: 59 AA.  
 AC R51402;  
 DT 26-OCT-1994 (first entry)  
 KW Branched peptide 3KH7.  
 KW Branch: lysine dimer; lysine octamer; immunocassay; reagent;  
 KW Non-A, Non-B hepatitis; NANBH; NANBH-related antibody;  
 KW hepatitis C; HCV; ELISA; passive haemagglutination; vaccine.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT modified site 59  
 FT /note="Linked to a lysine dimer to form branched  
 FT molecule"  
 PN W09406826-A.  
 PD 31-MAR-1994.  
 PE 15-SEP-1993; U08638.  
 PR 15-SEP-1992; US-946054.  
 PP (UNBI-) UNITED BIOMEDICAL INC.  
 PI Hosein B, Wang CY;  
 DR WPI: 94-118396/14.  
 PT New branched peptides for detecting non-A, non-B hepatitis antibodies  
 PT - to identify carriers and diagnose hepatitis C infection, contain at  
 PT least one specific epitope  
 PS Claim 24; Page 12; 43pp; English.  
 CC The sequences given in R51389-426 are peptides which form branched  
 CC peptides through either a lysine dimer or a lysine octamer at the C-  
 CC terminus. These branched peptides are immunocassay reagents for  
 CC detecting Non-A, Non-B hepatitis (NANBH)-related antibodies. They  
 CC may be used in the diagnosis of hepatitis C (HCV) and for the  
 CC identification of carriers, particularly by ELISA or passive  
 CC haemagglutination. They can also be used as vaccines to protect  
 CC against HCV infection.  
 SQ Sequence 59 AA.

Query Match 100.0%; Score 19; DB 10; Length 59;  
 Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 31 pyldgmm 38  
 QY 2 PXXXXXXM 9

RESULT 4  
 ID R39794 standard; protein: 60 AA.  
 AC R39794;  
 DT 11-JAN-1994 (first entry)  
 DE Human Kunitz-type protease inhibitor variant #1.  
 KW Amplify; primer; human; placenta; kunitz-type; protease; inhibitor;  
 KW domain; HK189; probe; pancreatitis; inflammation; thrombocytopenia;  
 OS Synthetic.  
 PN W09314123-A.  
 PD 22-JUL-1993.  
 PE 07-JAN-1993; DK0006.  
 PR 07-JAN-1992; WO-DK0003.  
 PP (NOVO) NOVO-NORDISK AS.  
 PI Bjorn SE, Foster DC, Norris F, Olsen OH;  
 DR WPI: 93-243151/30.  
 PT New human Kunitz-type protease inhibitor and variants - are for  
 PT treating pathological proteolysis, e.g. pancreatitis,  
 PT inflammation or thrombocytopenia, and derived DNA, vectors and  
 PT host cells  
 PS Disclosure; Page 27; 38pp; English.  
 CC The sequences given in R39794-96 represent variants based on the human  
 CC kunitz-type protease inhibitor domain HK189. The DNA encoding  
 CC the wild type protein domain was isolated from human placental DNA  
 CC by polymerase chain reaction (PCR) and mutated using mutagenic primers  
 CC in PCR. These variant inhibitors and the wild type protein (see also  
 CC R39792) may be used in the treatment of acute pancreatitis,  
 CC inflammation, thrombocytopenia or other conditions involving  
 CC pathological proteolysis.  
 SQ Sequence 60 AA.

Query Match 100.0%; Score 19; DB 8; Length 60;  
 Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 4 pnvcafp 11  
 QY 2 PXXXXXXM 9

RESULT 5  
 ID R39792 standard; protein: 60 AA.  
 AC R39792;  
 DT 11-JAN-1994 (first entry)  
 DE Human Kunitz-type protease inhibitor.  
 KW Amplify; primer; human; placenta; kunitz-type; protease; inhibitor;  
 KW domain; HK189; probe.  
 OS Homo sapiens.  
 PN W09314123-A.  
 PD 22-JUL-1993.  
 PE 07-JAN-1993; DK0006.  
 PR 07-JAN-1992; WO-DK0003.  
 PP (NOVO) NOVO-NORDISK AS.  
 PI Bjorn SE, Foster DC, Norris F, Olsen OH;  
 DR WPI: 93-243151/30.  
 PT New human Kunitz-type protease inhibitor and variants - are for  
 PT treating pathological proteolysis, e.g. pancreatitis,  
 PT inflammation or thrombocytopenia, and derived DNA, vectors and  
 PT host cells  
 PS Claim 1; Page 31; 38pp; English.  
 CC This sequence represents a human Kunitz-type protease inhibitor  
 CC domain. The DNA encoding this protein was isolated from human  
 CC placental DNA by polymerase chain reaction. This inhibitor and its  
 CC variants (see also R39793) may be used in the treatment of acute  
 CC pancreatitis, inflammation, thrombocytopenia or other conditions

CC Involving pathological proteolysis.  
SQ Sequence 60 AA;

Query Match 100.0%; Score 19; DB 8; Length 60;  
Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 4 pncvafpm 11  
QY 2 PXXXXXXM 9

RESULT 6  
ID R33871 standard; peptide: 67 AA.

AC R33871;  
DE 19-JUL-1993 (first entry)  
DR Polypeptide p1684 comprising HCV viral antigen.  
KW Hepatitis C virus; NANBH; assay; antibody; p380-JH1; p380-J; p380LG;  
P408.

OS Synthetic.  
PN WO9306247-A.  
PD 01-APR-1993.  
PF 16-SEP-1992; 007813.  
PR 16-SEP-1991; US-760292.  
PA (ABRO) ABBOTT LAB.  
PI Lesniowski R, Leung TK.  
DR WPI: 93-117563/14.

PT Assay for detecting presence of antibody to hepatitis C viral antigen - by contacting sample with polypeptide contg. at least one epitope of virus antigen  
PT Disclosure: Page 13; 63pp; English.

CC The synthetic peptide p1684 represents amino acid residues 1684-1750 of the hepatitis C viral antigen. The peptide may be used in an assay to detect antibodies to HCV and thus to diagnose chronic HCV infection.  
CC See also R33861-87.

CC Sequence 67 AA;

Query Match 100.0%; Score 19; DB 7; Length 67;  
Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 33 pyleggmm 40  
QY 2 PXXXXXXM 9

RESULT 7  
ID R32028 standard; Protein: 110 AA.

AC R32028;  
DE 05-JUL-1993 (first entry)  
DR Variant 19E - mutant EmuT 61.  
KW High affinity; FCEH; low affinity; FCEH; Padlan;  
OS IGE receptor; Fc; 19G1.  
OS Homo sapiens.

FT Key Location/Qualifiers  
FT region 7..12

FT region /label= beta-strand\_A

FT region /label= loop\_AB

FT region /label= beta-strand\_B

FT region /label= loop\_BC

FT region /label= beta-strand\_C

FT region /label= loop\_CD

FT region /label= beta-strand\_D

FT region /label= loop\_DE

FT region /label= beta-strand\_E

FT region 79..86  
FT /label= loop\_EF  
FT region 87..94  
FT /label= beta-strand\_F  
FT region 95..100  
FT /label= loop\_FG  
FT region 101..105  
FT /label= beta-strand\_G  
FT misc\_difference 49  
FT /label= mutation  
FT /note= "S -> A"

WO9304173-A.  
PN 04-MAR-1993.

PF 14-AUG-1992; 006860.  
PR 14-AUG-1991; US-744768.  
PR 07-MAY-1992; US-879495.  
PA (GETH) GENENTECH INC.  
PI Jardiou PM, Presta LG.  
DR WPI: 93-094004/11.

PT Polypeptide(s) binding to specific Fc epsilon receptors - act as Fc epsilon antagonists; useful for treating and preventing IGE-mediated disorders e.g. allergies  
PT Disclosure: Page 73; 113pp; English.  
CC IGE mutants were prepd. to evaluate their effect on binding to anti-IGE, esp. Me11, and to Fc epsilon RI and Fc epsilon RII.  
CC Some of the mutants were designed to substitute for a specific amino acid residue another residue with either similar or very different charge or size.  
CC Mutant 61 shows +ve binding to Fc epsilon RI and Fc epsilon RII.  
CC Sequence 110 AA;

Query Match 100.0%; Score 19; DB 6; Length 110;  
Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 94 phlpralm 101  
QY 2 PXXXXXXM 9

RESULT 8  
ID P90149 standard; protein: 115 AA.

AC P90149;  
DE 1-NOV-1989 (first entry)  
DR Sequence of hepatitis C virus cDNA insert in clone 7e  
KW Hepatitis C virus; clone 8h; clone 7e; probe; vaccine.  
OS Pan troglodytes

FT Key Location/Qualifiers  
FT region 101..115

PN GB2212511-A.

PN 26-JUL-1989.

PF 18-NOV-1988; G27024.

PF 18-NOV-1987; US-122714.

PA (CHIR) Chiron Corporation.

PI Houghton M, Choo QL, Kuo G;

DR WPI: 89-215054/30.

DR N-PSDB; N90318.

PT Hepatitis C virus gene - used for prodn. of polynucleotide probes, polypeptide(s) and antibodies for diagnosis, prevention and treatment of infection.

PT Disclosure: fig 17; 235pp; English.

CC The sequence is the peptide encoded by the hepatitis C virus (HCV) cDNA insert in clone 7e (see N90318). The polypeptides are used to diagnose HCV-induced NANBH, to raise antibodies for CC immunassay or treatment, or to produce vaccines.

CC The region shown overlaps with clone 8h.  
SQ Sequence 115 AA;

Query Match 100.0%; Score 19; DB 1; Length 115;  
Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 72 pkgpv1qm 79

OY 2 PXXXXXXM 9

# RESULT 9

ID R65162 standard; Protein: 121 AA.

DE 19-JAN-1996 (first entry)

KW Humanised anti-PR1A3 antibody VH chain.

KW Primer: amplify; polymerase chain reaction; PCR; human;

KW Biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;

KW CEA; chimeric protein; PR1A3 epitope; anti-PR1A3 antibody;

KW Colorectal carcinoma; monoclonal antibody.

OS Homo sapiens.

OS Mus musculus.

FT Key Location/Qualifiers

FT Peptide 31..35

FT /label- CDR1

FT Peptide 50..66

FT /label- CDR2

FT Peptide 99..110

FT /label- CDR3

FT WO9506067-A1.

PD 02-MAR-1995.

PR 19-AUG-1994; G01816.

PR 21-AUG-1993; GB-017423.

PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.

PI Bates PA, Bodmer WF, Durbin H, Snary D, Stewart LMD;

PI Young S;

DR WPI; 95-106813/14.

PT New molecules which bind carcinoembryonic antigen - used for the

PT diagnosis and treatment of colorectal carcinoma and for isolation

PT and purifications.

PS Disclosure; Fig.1; 67pp; English.

CC This sequence represents the VH chain of a humanised antibody which

CC comprises the framework regions of RF-TS3/CL and the complementarity

CC determining regions of the murine PR1A3 antibody given in R65155-56.

CC The murine antibody was raised by immunising a mouse with a hybrid

CC protein containing the B3-GPI region of human membrane-bound

CC carcinoembryonic antigen (CEA) joined to the C-terminal of a fragment

CC of biliary glycoprotein (BGP). The antisera produced was then absorbed

CC with purified BGP to remove BGP-reactive antibodies to leave CEA-

CC reactive antibodies. The complementarity determining regions of the

CC murine antibody are then inserted into this human backbone by PCR to

CC produce this humanised antibody. The antibody of the invention is

CC raised against a specific antigen which is part of human CEA. The

CC PR1A3 epitope was found to be an epitope of CEA within the B3-GPI

CC region. Antibodies which recognise the PR1A3 epitope are used in the

CC detection of well and poorly differentiated colorectal carcinomas. The

CC isolation of the specific PR1A3 epitope allows the development of

CC monoclonal antibodies specific for colorectal carcinoma. They can be

CC used in the study, isolation and purification of molecules to which

CC they specifically bind and the imaging and treatment of cells

CC exhibiting the molecules.

CC Sequence 121 AA;

SQ

Query Match 100.0%; Score 19; DB 14; Length 121;

Best Local Similarity 25.0%; Pred. No. 1.44e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 41 pygglewm 48

OY 2 PXXXXXXM 9

RESULT 10

ID R92515 standard; Protein: 123 AA.

AC R92515;

DE 09-MAY-1996 (first entry)

KW AR-2332 ORF 6 product (nucleocapsid protein).

KW Arterivirus; porcine reproductive and respiratory syndrome; PRRS;

KW vaccine; genetic immunization; diagnosis; vector.

OS Arteriviridae (unclassified) strain ATCC VR-2332.

PN WO9604010-A1.

PD 15-FEB-1996.

PF 04-AUG-1995; U09927.

PR 05-AUG-1994; US-287941.

PA (MINU ) UNIV MINNESOTA.

PI Elam MR, Kakach LT, Murtaugh MP;

DR WPI; 96-129128/13.

DR N-PSDB; T16245.

PT Viral DNA from VR-2332 genome ORF 2-ORF 7 region - causes porcine

PT reproductive and respiratory syndrome, useful in vaccines

PS Example 7; Page 46-47; 90pp; English.

CC Viral proteins (R92510-15) encoded by open reading frames ORF2,

CC ORF3, ORF4, ORF5, ORF6 and ORF7 (T16240-45) of the VR-2332 strain

CC have use in vaccine prodn. They are obtd. by expression of the

CC ORFs (singly or in combination) in prokaryotic or eucaryotic host

CC cells. The ORF7 product appears to form the nucleocapsid protein

CC in PRRS virus. It shares 64% identity with the ORF7 product of

CC the Lelystad strain of PRRS.

CC Sequence 123 AA;

SQ

Query Match 100.0%; Score 19; DB 16; Length 123;

Best Local Similarity 25.0%; Pred. No. 1.44e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 18 pynglcm 25

OY 2 PXXXXXXM 9

RESULT 11

ID R72829 standard; Protein: 158 AA.

AC R72829;

DE 23-OCT-1995 (first entry)

DE Human TIABP1.

KW TIABP1; TIA-1 binding protein; lymphocyte-mediated cytolytic;

KW cytotoxic; antitumor; cell death; apoptosis; cancer; therapy.

OS Homo sapiens.

PN WO9509924-A.

PD 13-APR-1995.

PF 07-OCT-1994; U11053.

PR 07-OCT-1993; US-133530.

PA (DAND ) DANA FARBER CANCER INST INC.

PI Anderson PJ, Tian Q;

DR WPI; 95-15265/20.

DR N-PSDB; Q86859.

PT TIA-1 binding proteins and isolated cDNA encoding them -bind to

PT cytoxic granule-associated RNA binding proteins and prevent

PT target cell death

PS Claim 35; Page 38; 100pp; English.

CC cDNAs encoding TIA-1 binding proteins (TIABPs) were identified by co-

CC transforming yeast GGT::171 cells with pMA424(TIA-1) and clones from

CC a library of human B-cells in which cDNAs were expressed as fusion

CC proteins having the GAL4 activation domain at the N-terminus.

CC Selection for double transformants expressing beta-galactosidase

CC identified the cDNA given in Q86859 encoding TIABP1. TIABP1

CC interacts with the RNA-binding domain of TIA-1, (in)directly

CC triggering programmed cell death.

SQ

Query Match 100.0%; Score 19; DB 13; Length 158;

Best Local Similarity 25.0%; Pred. No. 1.44e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 32 pdgtmilm 39

OY 2 PXXXXXXM 9

RESULT 12

ID R51467 standard; Protein: 172 AA.

AC R51467;

DE 15-NOV-1994 (first entry)

DE AK beta subunit R30A.  
 KW Mutant; C. glutamicum; aspartokinase; AK; alpha; beta; subunit;  
 KW substitution; non-acidic; amino acid; Ala; additional; enzyme;  
 KW feedback inhibition; synergic inhibition; heat stability.  
 OS Corynebacterium glutamicum.  
 PN 106062866-A.  
 PD 08-MAR-1994.  
 PF 27-APR-1993; 101450.  
 PR 28-APR-1992; JP-110292.  
 PA (AJIN ) AJINOMOTO KK.  
 DR WPI: 94-121127/15.  
 FT P-PSDB; Q61601.  
 PT The mutant aspartokinase gene - used to produce AK which is  
 removed from feedback inhibition  
 PS Disclosure: Page 23; 28pp; Japanese.  
 CC This sequence represents a mutant version of the C. glutamicum asparto-  
 kinase (AK) beta subunit. The mutation is at position 30 and  
 corresponds to the substitution of a non-acidic amino acid, pref.  
 CC Ala. This mutation causes removal of the enzyme from additional  
 CC feedback inhibition. Inhibition by Thr is removed completely and  
 CC the synergic inhibition of Lys and Thr is also removed. The new AK  
 CC has increased heat stability.  
 SO Sequence 172 AA;

Query Match 100.0%; Score 19; DB 10; Length 172;  
 Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 109 pvtlaefm 116

OY 2 PXXXXXXM 9

RESULT 13  
 ID R54659 standard; Protein: 199 AA.  
 AC R54659;  
 DT 31-OCT-1994 (first entry)  
 DE Human CD69.  
 KW Immune system; thrombocyte development; signal transduction; probe;  
 KW assay; diagnosis; therapy.  
 OS Homo sapiens.  
 FT Key Location/Qualifiers  
 FT protein 79..199  
 FT /note="soluble polypeptide"  
 PN MO9410188-A.  
 PD 11-MAY-1994.  
 PF 28-OCT-1993; U10418.  
 PR 29-OCT-1992; US-971097.  
 PA (IMNV ) IMMUNEX CORP.  
 PI Hjerrild KA, Ziegler SR;  
 DR WPI: 94-167377/20.  
 DR N-PSDB; Q65340.  
 PT CD69 nucleic acids and polypeptide - used in the diagnosis,  
 PT therapy and study of the activation and regulation of the immune  
 PT system  
 PS Claim 10: Page 27; 35pp; English.  
 CC The sequence is that of human CD69. The CD69 protein is active in  
 CC the regulation and function of the immune system. The protein may  
 CC be used for blocking thymocyte development in vitro systems. The  
 CC soluble polypeptide can be used to competitively bind the ligand in vivo  
 CC thus inhibiting signal transduction activity via endogenous cell surface  
 CC bound CD69. CD69 may also be used to generate antibodies.  
 CC See also R54660.  
 SO Sequence 199 AA;

Query Match 100.0%; Score 19; DB 10; Length 199;  
 Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 69 ppgytlsm 76

OY 2 PXXXXXXM 9

RESULT 14  
 ID R41262 standard; peptide: 242 AA.  
 AC R41262;  
 DT 24-FEB-1994 (first entry)  
 DE VWF fragment Thr92-Val173.  
 KW von Willebrand; factor; VWF; inhibition; platelets; disulphide bonding;  
 KW platelet membrane glycoprotein; Ib-IX receptor; GPIb(a); thrombosis;  
 KW subendothelium; collagen; proteoglycan; 52/48; von Willebrand disease.  
 OS Homo sapiens.  
 FT Key Location/Qualifiers  
 FT disulfide\_bond 18..204  
 FT /note="Disulphide bond between Cys509-Cys695"  
 PN MO9316709-A.  
 PD 02-SEP-1993.  
 PF 23-FEB-1993; U02034.  
 PR 26-FEB-1992; US-841591.  
 PA (SCRI ) SCRIPS RES INST.  
 PI Ruggeri ZM, Ware JL;  
 DR WPI: 93-288111/36.

PT New polypeptide inhibiting binding of von Willebrand factor to  
 platelets - esp. for treating or preventing thrombosis, also  
 PT related nucleic acid, vectors, transformants and antibodies  
 PS Claim 6; Fig 1; 12pp; English.  
 CC The sequences given in R41254-67 are fragments of mature von  
 CC Willebrand factor (VWF) which may be used to inhibit binding of VWF  
 CC to platelets. These fragments are derived from the VWF domain which  
 CC binds to the platelet membrane glycoprotein Ib-IX receptor (GPIb(a)).  
 CC This fragment also contains binding domains for components of the  
 CC subendothelium, such as collagen and proteoglycans, however other  
 CC regions of VWF may be more important in these interactions. The  
 CC entire GPIb(a) binding region, designated 52/48 due to its molecular  
 CC weight, has been demonstrated to competitively inhibit the binding  
 CC of VWF to platelets. However, manipulation of the 52/48 fragment  
 CC or its unglycosylated 38 kD equivalent requires that the number of  
 CC cysteine residues is reduced and permanently alkylated. This reduces  
 CC the amount of disulphide bonding. Disulphide bonding causes the  
 CC formation of insoluble and biologically inactive polypeptide  
 CC aggregates unsuitable for use as therapeutics. All the cysteine  
 CC residues within the 52/48 VWF fragment naturally take part in  
 CC disulphide bonding although it is not known whether it is intra- or  
 CC interchain binding. These 52/48 fragment subfragments may be used  
 CC in the treatment of thrombosis and von Willebrand diseases.  
 SO Sequence 242 AA;

Query Match 100.0%; Score 19; DB 8; Length 242;  
 Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 212 pttlpphm 219

OY 2 PXXXXXXM 9

RESULT 15  
 ID R85251 standard; Protein: 293 AA.  
 AC R85251;  
 DT 24-JAN-1996 (first entry)  
 DE Pseudomonas mendocina SD702 lipase mutant Leu25.  
 KW Lipase; SD702; detergents; food processing; papermaking;  
 KW mutant Leu25; modified properties.  
 OS Pseudomonas mendocina.  
 PN WO9514783-A1.  
 PD 01-JUN-1995.  
 PF 21-NOV-1994; J01965.  
 PR 24-NOV-1993; JP-293631.  
 PA (SHOWA ) SHOWA DENKO KK.  
 PI Miyota Y, Onno K, Sasuga J, Yoneda T;  
 DR WPI: 95-206940/27.  
 PT Lipase gene from Pseudomonas and variants of it - for production of  
 PT recombinant lipase LP with appropriate properties for industrial use  
 PS Claim 9; Pages 17-18; 30pp; Japanese.  
 CC R85249-R85284 are mutant P. mendocina SD702 lipases. The mutants  
 CC have modified properties compared to the wild type lipase, and

Sun Sep 13 10:56:38 1998

US-08-452-843-30.rag

Page 6

CC can be used in detergent, food processing and papermaking fields.  
SQ Sequence 293 AA;

Query Match 100.0%; Score 19; DB 14; Length 293;  
Best Local Similarity 25.0%; Pred. No. 1.44e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 14 p1v1ghm 21  
QY 2 PXXXXXXM 9

Search completed: Fri Sep 11 13:59:43 1998  
Job time : 16 secs.



REFERENCE A57434  
#authors Sasaki, H.; Nagura, K.; Ishino, M.; Tobioke, H.; Kotani, K.;  
#journal J. Biol. Chem. (1995) 270:21206-21219  
#title Cloning and characterization of cell adhesion kinase beta, a novel protein-tyrosine kinase of the focal adhesion kinase subfamily.  
#accession B57434  
#status preliminary  
#molecule\_type mRNA  
#residues 1-150 ##label SAS  
##cross-references G6:ID5853; NID:91000676; PID:d100884; PID:91000677  
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein  
KEYWORDS kinases; protein kinase homology  
SUMMARY phosphotransferase  
#length 150 #checksum 4845

Db 19 PEGPGEPM 26  
QY 2 PXXXXXXM 9

RESULT 3  
ENTRY A38956 #type complete  
TITLE IGF-dependent histamine-releasing factor - rabbit  
ALTERNATE\_NAMES 21k tumor protein; tumor-associated protein  
ORGANISM #formal\_name Oryctolagus cuniculus #common\_name domestic rabbit  
DATE 11-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Sep-1997

ACCESSIONS A38956  
REFERENCE A38956  
#authors Dawson, S.P.  
#submission submitted to GenBank, November 1994  
#description A rabbit gene encoding a protein homologous to a translationally controlled human tumour protein undergoes developmental regulation during mammary gland development.  
#accession A38956  
#status preliminary  
##molecule\_type mRNA  
##residues 1-172 ##label DAM  
##cross-references EMBL:Z46805; NID:9599941; PID:9599942  
CLASSIFICATION #superfamily IGF-dependent histamine-releasing factor lymphocyte  
KEYWORDS #length 172 #molecular\_weight 19425 #checksum 3009  
SUMMARY

Query Match 100.0%; Score 19; DB 2; Length 172;  
Best Local Similarity 25.0%; Pred. No. 2.29e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 108 PERVKPFM 115  
QY 2 PXXXXXXM 9

RESULT 4  
ENTRY H69644 #type complete  
TITLE Initiation factor IF-3 infC - Bacillus subtilis  
ORGANISM #formal\_name Bacillus subtilis  
DATE 03-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 26-Feb-1998

ACCESSIONS H69644  
REFERENCE H69644  
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertolo, M.G.; Bessières, P.; Boloitin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Brann, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.;

Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fume, S.; Gallizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golligly, E.J.; Grandi, G.; Guseppl, G.; Guy, B.J.; Haga, K.; Hahle, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Kleerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koeltter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetle, D.; Porwolk, S.; Prescott, A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.O.; Serrro, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Taccini, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wandut, R.; Wedler, E.; Wedler, H.; Weltzenegger, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, Y.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;  
#journal #title Nature (1997) 390:249-256  
#accession H69644  
#status preliminary; nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA  
#residues 1-173 ##label KUN  
##experimental\_source strain 168  
GENETICS  
#gene infC  
CLASSIFICATION #superfamily translation initiation factor IF-3  
SUMMARY #length 173 #molecular\_weight 19732 #checksum 2183

Query Match 100.0%; Score 19; DB 2; Length 173;  
Best Local Similarity 25.0%; Pred. No. 2.29e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 155 PKMDGRSM 162  
QY 2 PXXXXXXM 9

RESULT 5  
ENTRY CYCAG1 #type complete  
TITLE gamma-crystallin ml - common carp  
ORGANISM #formal\_name Cyprinus carpio #common\_name common carp  
DATE 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 05-Sep-1997

ACCESSIONS JS0061  
REFERENCE JS0061  
#authors Chang, T.; Jiang, Y.J.; Chlou, S.H.; Chang, W.C.  
#journal Blochm. Biophys. Acta (1988) 951:226-229  
#title Carp gamma-crystallins with high methionine content: cloning and sequencing of the complementary DNA.  
#cross-references MUID:89051015  
#accession JS0061  
##molecule\_type mRNA  
##residues 1-178 ##label CHA  
##cross-references EMBL:X12902; NID:962601; PID:962602; EMBL:M33115  
#note the authors translated the codon ACC for residue 101 as Asn  
CLASSIFICATION #superfamily beta-crystallin  
KEYWORDS duplication; eye lens



FEATURE  
2-178 #product gamma-crystallin m1 #status predicted #label  
MAY

2-40 #domain crystallin repeat #label GK1\  
41-86 #domain crystallin repeat #label GK2\  
92-132 #domain crystallin repeat #label GK3\  
133-172 #domain crystallin repeat #label GK4  
SUMMARY #length 178 #molecular-weight 21882 #checksum 5383

Query Match 100.0%; Score 19; DB 1; Length 178;  
Best Local Similarity 25.0%; Pred. No. 2.29e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 87 PYRGSYRN 94  
QY 2 PXXXXXXM 9

RESULT 6 JMB0 #type complete  
ENTRY ameloegenin I precursor - bovine  
TITLE leucine-rich amelogenin peptide, LRAP-43; leucine-rich  
CONTAINS amelogenin peptide, LRAP-48; leucine-rich amelogenin  
peptide, LRAP-59; tyrosine-rich amelogenin peptide,  
TRAP-43; tyrosine-rich amelogenin peptide, TRAP-45  
#formal name Bos primigenius taurus #common name cattle  
28-May-1986 #sequence\_revision 06-Feb-1995 #text\_change  
05-Sep-1997

ACCESSIONS A37998; JN0123; A26549; A03300; PC2005; PC2007  
REFERENCE A37998

#authors Gibson, C.; Golub, E.; Herold, R.; Riser, M.; Ding, W.;  
Shimokawa, H.; Young, M.; Termini, J.; Rosenbloom, J.  
#journal Biochemistry (1991) 30:1075-1079  
#title Structure and expression of the bovine amelogenin gene.  
#cross-references M01D:91113686  
#accession A37998  
#molecule-type mRNA  
#residues 1-213 #label G12  
#cross-references GB:M63499; NID:g162659; PID:g162660  
JN0123

REFERENCE JN0123

#authors Gibson, C.W.; Golub, E.; Ding, W.; Shimokawa, H.; Young, M.;  
Termini, J.; Rosenbloom, J.  
#journal Biochem. Biophys. Res. Commun. (1991) 174:1306-1312  
#title Identification of the leucine-rich amelogenin peptide (LRAP)  
as the translation product of an alternatively spliced  
transcript.  
#cross-references M01D:91144612  
#accession JN0123  
#molecule-type mRNA  
#residues 7-49,188-211 #label G18  
#cross-references GB:M63631; NID:g163308; PID:g163309  
#experimental-source tooth  
A26549

REFERENCE A26549

#authors Shimokawa, H.; Sobel, M.E.; Sasaki, M.; Termini, J.D.; Young,  
M.F.  
#journal J. Biol. Chem. (1987) 262:4042-4047  
#title Heterogeneity of amelogenin mRNA in the bovine tooth germ.  
#cross-references M01D:87166009  
#accession A26549  
#molecule-type mRNA  
#residues 49-65,'Y',67-108,116-143,153-179,'V',181-183,'L',  
185-188,'M',190-192,'I',194-213 #label SHI  
#cross-references GB:302696  
A03300

REFERENCE A03300

#authors Takagi, T.; Suzuki, M.; Baba, T.; Minegishi, K.; Sasaki, S.  
#journal Biochem. Biophys. Res. Commun. (1984) 121:592-597  
#title Complete amino acid sequence of amelogenin in developing  
bovine enamel.  
#cross-references M01D:84231410  
#accession A03300  
#molecule-type protein  
#residues 17-33,'Q',35-55,'Q',57-76,'D',78-82,'P',84-108,117-125,  
'T',127-152,'L',154-162,'H',164,'P',165-166,'I',168,

REFERENCE PC2005  
#authors Fincham, A.G.; Moradian-Oldak, J.  
#journal Biochem. Biophys. Res. Commun. (1993) 197:248-255  
#title Amelogenin post-translational modifications: carboxy-terminal  
processing and the phosphorylation of bovine and porcine  
"LRAP" and "LRAP" amelogenins.  
#accession PC2005  
#molecule-type protein  
#residues 17-61 #label FI2  
#accession PC2007  
#molecule-type protein  
#residues 17-40;48-49,188-202 #label FI1  
COMMENT Amelogenin is the predominant protein in developing dental enamel.  
GENETICS  
#map\_position X  
CLASSIFICATION #superfamily amelogenin  
KEYWORDS alternative splicing; enamel; phosphoprotein; tandem repeat;  
tooth

FEATURE  
1-16 #domain signal sequence #status predicted #label SIG\  
17-213 #product amelogenin #status predicted #label MAY\  
17-61 #product tyrosine-rich amelogenin peptide, TRAP-45  
17-59 #status experimental #label MA2\  
#product tyrosine-rich amelogenin peptide, TRAP-43  
17-49,188-213 #status experimental #label MA3\  
#product leucine-rich amelogenin peptide, LRAP-59  
17-49,188-202 #status experimental #label MA4\  
#product leucine-rich amelogenin peptide, LRAP-48  
17-49,188-196 #product leucine-rich amelogenin peptide, LRAP-43  
136-165 #status experimental #label MA6\  
#region 3-residue repeats (Q-P-X)  
32 #binding site phosphate (Ser) (covalent) #status  
predicted

SUMMARY #length 213 #molecular-weight 24119 #checksum 90

Query Match 100.0%; Score 19; DB 1; Length 213;  
Best Local Similarity 25.0%; Pred. No. 2.29e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 38 PLKWTQSM 45  
QY 2 PXXXXXXM 9

RESULT 7 S22517 #type complete  
ENTRY S-allele-associated glycoprotein So precursor - garden  
TITLE petunia  
ORGANISM #formal\_name Petunia x hybrida #common\_name garden petunia  
DATE 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change  
08-Sep-1997  
S22517; S18611

ACCESSIONS S22517  
REFERENCE S22516

#authors Al, Y.; Tsal, D.S.; Kao, T.  
#journal Plant Mol. Biol. (1992) 19:523-528  
#title Cloning and sequencing of cDNAs encoding two S proteins of a  
self-compatible cultivar of Petunia hybrida.  
#accession S22517  
#molecule-type mRNA  
#residues 1-222 #label A1Y  
#cross-references EMBL:M81686; NID:g169249; PID:g169250  
the authors translated the codon TAT for residue 110 as  
Thr  
S18610

REFERENCE S18610

#authors Al, Y.; Kiron, E.; Kao, T.  
#journal Mol. Gen. Genet. (1991) 230:353-358  
#title S-alleles are retained and expressed in a self-compatible  
cultivar of Petunia hybrida.  
#accession S18611

```
##molecule_type protein
##residues 23-39 ##label AIR
CLASSIFICATION #superfamily Enterobacter ribonuclease
KEYWORDS glycoprotein
FEATURE
1-22
23-222
#domain signal sequence #status experimental #label SIG\
#product S-allele-associated protein So #status
experimental #label MAT
SUMMARY #length 222 #molecular-weight 25432 #checksum 4481
Query Match 100.0%; Score 19; DB 2; Length 222;
Best Local Similarity 25.0%; Pred. No. 2.29e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 59 PDSVSYM 66
OY 2 PXXXXXX 9
RESULT 8
ENTRY S53619 #type complete
TITLE major prion protein - Presbytis francisci
ORGANISM #formal_name Presbytis francisci
DATE 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change
08-Sep-1997
ACCESSIONS S53619; S71057
REFERENCE S53614
#authors Schatzl, H.M.; da Costa, M.; Taylor, L.; Cohen, F.E.;
#journal J. Mol. Biol. (1995) 245:362-374
#title Prion protein gene variation among primates.
#accession S53619
#status nucleic acid sequence not shown
#molecule_type DNA
#residues 1-253 ##label SCH
#cross-references EMBL:U08302
REFERENCE S71041
#authors Schatzl, H.M.
#submission submitted to the EMBL Data Library, April 1994
#accession S71057
#molecule_type DNA
#residues 1-210, 'E', 212-253 ##label SCM
#cross-references EMBL:U08302; NID:G1396067; PID:G1396068
CLASSIFICATION #superfamily major prion protein
KEYWORDS amyloid; brain; glycoprotein; lipoprotein; prion; scrapie;
transmembrane protein
SUMMARY #length 253 #molecular-weight 27687 #checksum 917
Query Match 100.0%; Score 19; DB 2; Length 253;
Best Local Similarity 25.0%; Pred. No. 2.29e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 102 PSKPTSM 109
OY 2 PXXXXXX 9
RESULT 9
ENTRY A32057 #type complete
TITLE anti protein - Azotobacter vinelandii
ORGANISM #formal_name Azotobacter vinelandii
DATE 20-Jul-1989 #sequence_revision 20-Jul-1989 #text_change
20-Mar-1998
ACCESSIONS A32057
REFERENCE A32057
#authors Joergers, R.D.; Jacobson, M.R.; Premakumar, R.; Wolfinger,
E.D.; Bishop, P.E.
#journal J. Bacteriol. (1989) 171:1075-1086
#title Nucleotide sequence and mutational analysis of the structural
genes (antiHdk) for the second alternative nitrogenase from
Azotobacter vinelandii.
#cross-references MUID:89123105
#accession A32057
```

```
##status preliminary
##molecule_type DNA
##residues 1-275 ##label JOE
#cross-references GB:M23528; NID:G142378; PID:G142379
CLASSIFICATION #superfamily nitrogenase iron protein
KEYWORDS ATP; iron-sulfur protein
FEATURE
9-16
15 #region nucleotide-binding motif A (P-loop)\
40,44 #binding-site Asp (Lys) #status predicted\
97,132 #active-site Asp #status predicted\
#binding-site 4Fe-4S cluster (Cys) (covalent) (shared
with dimeric partner) #status predicted
SUMMARY #length 275 #molecular-weight 29885 #checksum 98
Query Match 100.0%; Score 19; DB 2; Length 275;
Best Local Similarity 25.0%; Pred. No. 2.29e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 54 PEETLDM 61
OY 2 PXXXXXX 9
RESULT 10
ENTRY A53570 #type complete
TITLE collectin-43 - bovine
ALTERNATE_NAMES lectin CL-43
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 12-Apr-1995 #sequence_revision 23-Feb-1996 #text_change
23-May-1997
ACCESSIONS A53570; A46689
REFERENCE A53570
#authors Lim, B.L.; Willis, A.C.; Reid, K.B.M.; Lu, J.; Laursen, S.B.;
Jensenius, J.C.; Holmskov, U.
#journal J. Biol. Chem. (1994) 269:11820-11824
#title Primary structure of bovine collectin-43 (CL-43). Comparison
with conglutinin and lung surfactant protein-D.
#accession A53570
#status preliminary
#molecule_type mRNA
#residues 1-301 ##label LIM
#cross-references GB:X75912
REFERENCE A46689
#authors Holmskov, U.; Telsner, B.; Willis, A.C.; Reid, K.B.;
Jensenius, J.C.
#journal J. Biol. Chem. (1993) 268:10120-10125
#title Purification and characterization of a bovine serum lectin
(CL-43) with structural homology to conglutinin and SP-D
and carbohydrate specificity similar to mannan-binding
protein.
#accession A46689
#molecule_type protein
#residues 1-27 ##label HOL
#experimental_source serum
#note sequence extracted from NCBI backbone (NCBIP-131234)
CLASSIFICATION #superfamily pulmonary surfactant protein D; C-type lectin
homology
KEYWORDS lectin
FEATURE
177-299
SUMMARY #domain C-type lectin homology #label LCH
#length 301 #molecular-weight 31419 #checksum 4993
Query Match 100.0%; Score 19; DB 2; Length 301;
Best Local Similarity 25.0%; Pred. No. 2.29e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 60 PSGRGSW 67
OY 2 PXXXXXX 9
RESULT 11
ENTRY A40943 #type complete
```

TITLE laponizing lipase (EC 3.1.1.-) precursor - Pseudomonas sp.  
 ALTERNATE\_NAMES lipase P  
 ORGANISM #formal\_name Pseudomonas sp.  
 DATE 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 31-Oct-1997

ACCESSIONS A40943; B40943  
 REFERENCE A40943  
 #authors Ihara, F.; Kageyama, Y.; Hirata, M.; Nihira, T.; Yamada, Y.  
 #journal J. Biol. Chem. (1991) 266:18135-18140  
 #title Purification, characterization, and molecular cloning of laponizing lipase from Pseudomonas species.

#cross-references M01D:92011544  
 #accession A40943  
 #molecule\_type DNA  
 #residues 1-311 #label IHA  
 #cross-references GB:D10166; GB:D90398; NID:g216901; PID:d1001504; PID:g216902

#accession B40943  
 #molecule\_type protein  
 #residues 27-45 #label IH2  
 COMMENT This extracellular lipase catalyzes the synthesis of macrocyclic lactones in anhydrous organic solvents.

CLASSIFICATION #superfamily Pseudomonas triacylglycerol lipase  
 KEYWORDS carboxylic ester hydrolase; hydrolase  
 FEATURE 1-26  
 27-311 #domain signal sequence #status experimental #label SIG\

SUMMARY #product laponizing lipase #status experimental #label NAT  
 #length 311 #molecular-weight 32737 #checksum 580

Query Match 100.0%; Score 19; DB 2; Length 311;  
 Best Local Similarity 25.0%; Pred. No. 2.29e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 35 PIVLAHGM 42  
 QY 2 PXXXXXXM 9

RESULT 12  
 ENTRY S25768 #type complete  
 TITLE triacylglycerol lipase (EC 3.1.1.3) precursor - Pseudomonas aeruginosa  
 ALTERNATE\_NAMES lipase A precursor  
 ORGANISM #formal\_name Pseudomonas aeruginosa  
 DATE 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 20-Mar-1998

ACCESSIONS S25768; S24160; S21751; JT0958  
 REFERENCE S25768  
 #authors Wohlfarth, S.; Hoessche, C.; Strunk, C.; Winkler, U.K.  
 #journal J. Gen. Microbiol. (1992) 138:1325-1335  
 #title Molecular genetics of the extracellular lipase of Pseudomonas aeruginosa PAO1.

#accession S25768  
 #molecule\_type DNA  
 #residues 1-311 #label WOH  
 #cross-references EMBL:X63390; NID:g45340; PID:g45341  
 REFERENCE S24160  
 #authors Chihara-Stomi, M.; Yoshikawa, K.; Oshima-Hirayama, N.; Yamamoto, K.; Sogabe, Y.; Nakatani, T.; Nishioke, T.; Oda, J.  
 #journal Arch. Biochem. Biophys. (1992) 296:505-513  
 #title Purification, molecular cloning, and expression of lipase from Pseudomonas aeruginosa.

#cross-references M01D:92337414  
 #accession S24160  
 #status preliminary  
 #molecule\_type DNA  
 #residues 1-155; 'I', 157-201, 'H', 203, 'V', 205-311 #label CHI  
 #cross-references GB:D10048; NID:g216895; PID:d1001402; PID:g216896  
 REFERENCE S21751  
 #authors Jaeger, K.E.; Adrian, F.J.; Meyer, H.E.; Hancock, R.E.W.; Winkler, U.K.

#journal Blochim. Biophys. Acta (1992) 1120:315-321  
 #title Extracellular lipase from Pseudomonas aeruginosa is an amphiphilic protein.  
 #cross-references M01D:92247813  
 #accession S21751  
 #status preliminary  
 #molecule\_type protein  
 #residues 27-39, 'X', 41-53, 'X', 55-57 #label JAE

GENETICS #gene lipA  
 CLASSIFICATION #superfamily Pseudomonas triacylglycerol lipase  
 KEYWORDS carboxylic ester hydrolase  
 FEATURE 1-26  
 27-311 #domain signal sequence #status predicted #label SIG\

SUMMARY #product lipase #status predicted #label IIP  
 #length 311 #molecular-weight 32723 #checksum 451

Query Match 100.0%; Score 19; DB 2; Length 311;  
 Best Local Similarity 25.0%; Pred. No. 2.29e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 35 PIVLAHGM 42  
 QY 2 PXXXXXXM 9

RESULT 13  
 ENTRY P4BPF6 #type complete  
 TITLE P4 protein - phage phi-6  
 ORGANISM #formal\_name phage phi-6  
 #note host Pseudomonas phage phi-6  
 DATE 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 21-Nov-1997

ACCESSIONS C29685  
 REFERENCE A93032  
 #authors Mindich, L.; Nemhauser, I.; Gottlieb, P.; Romantschuk, M.; Carton, J.; Frucht, S.; Strassman, J.; Bamford, D.H.; Kalkkainen, N.  
 #journal J. Virol. (1988) 62:1180-1185  
 #title Nucleotide sequence of the large double-stranded RNA segment of bacteriophage phi-6: genes specifying the viral replicase and transcriptase.

#cross-references M01D:88155752  
 #accession C29685  
 #molecule\_type genomic RNA  
 #residues 1-332 #label MIN

COMMENT The genome of this phage consists of three segments of double-stranded RNA; this protein is coded by the largest segment (L).

GENETICS #gene 4  
 #map\_position segment L  
 CLASSIFICATION #superfamily phage phi-6 P4 protein  
 KEYWORDS early protein  
 SUMMARY #length 332 #molecular-weight 35163 #checksum 6416

Query Match 100.0%; Score 19; DB 1; Length 332;  
 Best Local Similarity 25.0%; Pred. No. 2.29e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 280 PLAADTHM 287  
 QY 2 PXXXXXXM 9

RESULT 14  
 ENTRY A54805 #type complete  
 TITLE protein kinase Mpk2 (EC 2.7.1.-) - African clawed frog  
 ORGANISM #formal\_name Xenopus laevis #common\_name African clawed frog  
 DATE 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 19-Dec-1997  
 ACCESSIONS A54805

```

REFERENCE      A54805
#authors      Rouse, J.; Cohen, P.; Trigon, S.; Morange, M.;
               Alonso-Llamazares, A.; Zamanillo, D.; Hunt, T.; Nebreda,
               A.R.
#journal      Cell (1994) 78:1027-1037
#title        A novel kinase cascade triggered by stress and heat shock
               that stimulates MAPKAP kinase-2 and phosphorylation of the
               small heat shock proteins.
#accession    A54805
#status       preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues     1-361 #label ROU
#cross-references GB:X80751; NID:9557679; PID:9557680
CLASSIFICATION #superfamily kinase-related transforming protein; protein
               kinase homology
KEYWORDS       phosphotransferase
FEATURE        23-278 #domain protein kinase homology #label KIN\
31-39 #region protein kinase ATP-binding motif
SUMMARY        #length 361 #molecular-weight 41718 #checksum 8091

Query Match      100.0%; Score 19; DB 2; Length 361;
Best Local Similarity 25.0%; Pred. No. 2.29e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 192 PEIMLNMM 199
OY 2 PXXXXXXM 9

RESULT 15
ENTRY   S56151 #type fragment
TITLE   tubulin alpha chain - Spathidium sp. (fragment)
ORGANISM #formal_name Spathidium sp.
DATE     19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change
         08-Sep-1997
ACCESSIONS S56151
REFERENCE  S56148
#author    Tournebise, A.B.; Tsao, N.; Klobutcher, L.A.; Pearlman,
           R.E.; Adoutte, A.
#journal    EMBO J. (1995) 14:3262-3267
#title      Genetic code deviations in the ciliates: evidence for
           multiple and independent events.
#accession  S56151
#status     nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues   1-382 #label TOU
#cross-references EMBL:249848; NID:9861141; PID:9861142
#note       the nucleotide sequence was submitted to the EMBL Data
           Library, June 1995
CLASSIFICATION #superfamily tubulin
KEYWORDS       heterodimer; microtubule
SUMMARY        #length 382 #checksum 9637

Query Match      100.0%; Score 19; DB 2; Length 382;
Best Local Similarity 25.0%; Pred. No. 2.29e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 231 PYPRHFM 238
OY 2 PXXXXXXM 9

Search completed: Fri Sep 11 14:00:27 1998
Job time : 27 secs.

```

(TM)

Distribution rights by Oxford Molecular Ltd

not generated.

1 XXXXXXXX 5

PAM 15

69111 seqs, 2500

### Listing first 45 summaries

1:swiss1

Mean 12.212; Variance 5.420; scale 2.253

and is derived by analysis of the total score distribution.

## SUMMARIES

Accession	Length	DB	ID	Description	Pred. No.
U00000.0	70	1	CCCN_P5ESC	CHOLECYSTOKININ.	2.88e+02
U00000.0	284	1	BAN7_MOUSE	ERYTHROCYTE BAND 7 INT	2.88e+02
U00000.0	287	1	BAN7_HUMAN	ERYTHROCYTE BAND 7 INT	2.88e+02
U00000.0	294	1	G3P_ESCHE	GLYCERALDEHYDE 3-PHOSP	2.88e+02
U00000.0	303	1	G3P_KLEPN	GLYCERALDEHYDE 3-PHOSP	2.88e+02
U00000.0	308	1	BEL1_SFVL	BEL-1 PROTEIN.	2.88e+02
U00000.0	349	1	CCR4_RAT	SDF-1 RECEPTOR (STROMA	2.88e+02
U00000.0	359	1	CCR4_MOUSE	SDF-1 RECEPTOR (STROMA	2.88e+02
U00000.0	364	1	EBG2_BOVIN	PROBABLE G PROTEIN-COU	2.88e+02
U00000.0	375	1	ACT2_OXYNO	ACTIN, CYTOPLASMIC (AC	2.88e+02
U00000.0	375	1	ACT2_OXYNR	ACTIN, CYTOPLASMIC (AC	2.88e+02
U00000.0	377	1	ACT1_SOLIU	ACTIN 58.	2.88e+02
U00000.0	377	1	ACT1_SOLRV	ACTIN 58.	2.88e+02
U00000.0	391	1	EBL1_ADEMT	EARLY E1B 44 KD PROTEI	2.88e+02
U00000.0	409	1	EBL1_CYCME	ELONGATION FACTOR TU (	2.88e+02
U00000.0	421	1	ARAB_CORL	ASPARTOKINASE ALPHA AN	2.88e+02
U00000.0	446	1	LVS9_YEAST	SACCHAROPINE DEHYDROE	2.88e+02
U00000.0	451	1	CLUS_COTJA	CLUSTERIN PRECURSOR (5	2.88e+02
U00000.0	467	1	HSP3_CHICK	HEAT SHOCK FACTOR PROT	2.88e+02
U00000.0	476	1	ATPB_CYACA	ATP SYNTHASE BETA CHAI	2.88e+02
U00000.0	485	1	ATPB_CYARA	ATP SYNTHASE BETA CHAI	2.88e+02
U00000.0	493	1	ADBO_YEAST	NADPH:ADP-DEPENDENT OXID	2.88e+02
U00000.0	507	1	CBS_YEAST	CYSTATHIONINE BETA-SYN	2.88e+02

24	19	100.0	510	1	CG39_HUMAN	VASCULAR ATP-DITHIOPHOS	2.88e+02
25	19	100.0	512	1	NARH_ECOLI	RESPIRATORY NITRATE RE	2.88e+02
26	19	100.0	514	1	COX1_HORSE	CYTOCHROME C OXIDASE P	2.88e+02
27	19	100.0	527	1	HSF8_LYCEP	HEAT SHOCK FACTOR PROT	2.88e+02
28	19	100.0	535	1	COX1_HAWKI	CYTOCHROME C OXIDASE P	2.88e+02
29	19	100.0	550	1	CBS_HUMAN	CYSTATHIONINE BETA-SYN	2.88e+02
30	19	100.0	661	1	KML1_HUMAN	INTERFERON-REGULATED R	2.88e+02
31	19	100.0	695	1	FSH_DROME	HEAT SHOCK FACTOR PROT	2.88e+02
32	19	100.0	695	1	FSHR_MACFA	FOLLICLE STIMULATING H	2.88e+02
33	19	100.0	700	1	B1B_DROME	NEUROGENIC PROTEIN BIG	2.88e+02
34	19	100.0	734	1	G13B_DIDDI	CELL SURFACE GLYCOPRO	2.88e+02
35	19	100.0	790	1	LY14_YEAST	LEUCINE BIOSYNTHESIS RE	2.88e+02
36	19	100.0	848	1	DYN3_RAT	DYNAMIN 3 (DYNAMIN, TE	2.88e+02
37	19	100.0	855	1	CNRC_BOVIN	CONE CGMP-SPECIFIC 3'	2.88e+02
38	19	100.0	858	1	EP2_RAT	ELONGATION FACTOR 2 (E	2.88e+02
39	19	100.0	870	1	DYN2_RAT	DYNAMIN 2	2.88e+02
40	19	100.0	985	1	DPOL_NPOPO	DNA POLYMERASE (EC 2.7	2.88e+02
41	19	100.0	1125	1	NAP4_MOUSE	MICROTUBULE-ASSOCIATED	2.88e+02
42	19	100.0	1255	1	ERB2_HUMAN	ERBB2 RECEPTOR PROTEI	2.88e+02
43	19	100.0	1342	1	ERB3_HUMAN	ERBB3 RECEPTOR PROTEI	2.88e+02
44	19	100.0	3969	1	HRX_HUMAN	ZINC FINGER PROTEIN HR	2.88e+02
45	19	100.0	4566	1	DYH8_CHIRE	DYNEIN BETA CHAIN, FLA	2.88e+02

## ALIGNMENTS

ID	RESULT	STANDARD	PRT	70 AA.
AC	CKKN.PSESC			
DT	P80345;			
DT	01-OCT-1994 (REL. 30, CREATED)			
DT	01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)			
DE	CHOLECYSTOKININ.			
OS	PSEUDOMYS SCRIPIA (SLIDER TURTLE).			
OC	EUMAROTIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; REPTILIA;			
OC	ANAPSIDA.			
RN	(1)			
RP	SEQUENCE.			
RX	TISSUE=SMALL INTESTINE;			
RA	MEDLINE: 95010049.			
RA	JOHNSON A.H.;			
RL	EUR. J. BIOCHEM. 224:691-702(1994);			
CC	-1- FUNCTION: THIS PEPTIDE HORMONE INDUCES GALL BLADDER CONTRACTION			
CC	AND THE RELEASE OF PANCREATIC ENZYMES IN THE GUT. ITS FUNCTION			
CC	IN THE BRAIN IS NOT CLEAR.			
CC	-1- TURTLE BRAIN CONTAINS CCK-OCCTAPEPTIDE (CCK8) AND CCK7; WHEREAS			
CC	-1- THE GUT CONTAINS INTRACT CCK33, CCK40 AND CCK58.			
CC	-1- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.			
CC	PROSITE; PS00229; GASTRIN; 1.			
DR	HORMONE; CLEAVAGE ON PAIR OF BASIC RESIDUES; AMIDATION; SULFATATION			
KW	HORMONE; CLEAVAGE ON PAIR OF BASIC RESIDUES; AMIDATION; SULFATATION			
FT	PEPTIDE 13 70			
FT	PEPTIDE 31 70			
FT	PEPTIDE 38 70			
FT	PEPTIDE 63 70			
FT	PEPTIDE 64 70			
FT	MOD_RES 64 64			
FT	MOD_RES 70 70			
SO	SEQUENCE 70 AA; 7855 MW; 52B2A940 CRC32;			
Query Match	100.0%;	Score 19;	DB 1;	Length 70;
Best Local Similarity	25.0%;	Pred. No. 2	886+02;	
Matches 2;	Conservative	0;	Mismatches	6;
			Indels	0;
			Gaps	
DB	39 PTGRISM46			
OY	2 PXXXXXXM 9			
RESULT	2	STANDARD;	PRT;	284 AA.
AC	PS4116; 060744; 062455;			
DT	01-OCT-1996 (REL. 34, CREATED)			
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)			

```

DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DT ERTHRHOCTE BAND 7 INTEGRAL MEMBRANE PROTEIN (STOMATIN) (PROTEIN
DE 7.2B)
CN EPR7.2 OR EPR7.2.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA, METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC EUKHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-BONE MARROW;
RX SCHLEGEL W., UNPUBLISHED I., PROHASKA R.;
RN GENE 178:115-118(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J;
RX GALLAGHER P.G., ROMANA M., LIEMAN J.H., WARD D.C.;
RN UNPUBLISHED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96374834.
RA GALLAGHER P.G., TURETSKY T., MENTZER W.C.;
RL GENOMICS 34:410-412(1996).
CC -1- FUNCTION: THOUGHT TO REGULATE CATION CONDUCTANCE.
CC -1- SUBCELLULAR LOCATION: EXPOSED ON THE CYTOPLASMIC SURFACE OF THE
CC MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE BAND 7/MC-2 FAMILY.
CC DR EMBL; X91043; G975689; -.
DR DR EMBL; U17297; G972907; -.
DR DR EMBL; U50999; G1469524; ALT SRO.
DR DR EMBL; U50993; G1469524; JOINED.
DR DR EMBL; U50994; G1469524; JOINED.
DR DR EMBL; U50995; G1469524; JOINED.
DR DR EMBL; U50996; G1469524; JOINED.
DR DR EMBL; U50997; G1469524; JOINED.
DR DR EMBL; U50998; G1469524; JOINED.
DR DR MED; MGI:95403; EPR7.2.
DR DR PROSITE; PS01270; BAND_7; 1.
RW ERTHRHOCTE; TRANSMEMBRANE; PHOSPHORYLATION.
FT DOMAIN 32 52
FT TRANSMEM 53 284
FT CONFLICT 37 37 V -> A (IN REF. 2).
FT CONFLICT 40 40 I -> F (IN REF. 2).
FT CONFLICT 43 43 L -> I (IN REF. 2).
FT CONFLICT 91 91 F -> L (IN REF. 2).
FT CONFLICT 273 273 V -> I (IN REF. 1).
FT CONFLICT 283 283 N -> H (IN REF. 1).
SO SEQUENCE 284 AA; 31403 MW; 7D15BAFF CRC32;

Query Match 100.0%; Score 19; DB 1; Length 284;
Best Local Similarity 25.0%; Pred. No. 2,886+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 200 PVDGRAM 207
Cy 2 PXXXXXXM 9

RESULT 3
ID BAN7_HUMAN STANDARD: PRT: 287 AA.
AC P27105; Q14087; Q15609;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ERTHRHOCTE BAND 7 INTEGRAL MEMBRANE PROTEIN (STOMATIN) (PROTEIN
DE 7.2B).
GN EPR7.2 OR BND7.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA, METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC EUKHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE; 91355220.

```

RA HIEBL-DRECHMIED C.M., ENTLER B., GLOTFMANN C., MAURER-FOGY I.,  
RA STATATOWA C., PROHASKA R.;  
RA BIOCHIM. BIOPHYS. ACTA 1090:123-124(1991).  
[2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE: 96423038.  
RA UNFRIED I., ENTLER B., PROHASKA R.;  
RA GENOMICS 30:521-528(1995).  
[3]  
RN SEQUENCE FROM N.A.  
RP TISSUE-BONE MARROW;  
RC STEWART G.W., HEPWORTH-JONES B.J., KEEN J.N., DASH B.J.C.,  
RA ARBENT A.C., CASIMIR C.M.;  
RL SUBMITTED (DEC-1991) TO EMBL/GENBANK/DBJ DATA BANKS.  
[4]  
RN SEQUENCE FROM N.A.  
RP MEDLINE: 96064711.  
RA GALLAGHER P.G., FORGET B.G.;  
RA J. BIOL. CHEM. 270:26358-26363(1995).  
[5]  
RN SEQUENCE OF 4-24, AND PHOSPHORYLATION SITE.  
RX MEDLINE: 93385136.  
RX SALZER U., AHORN H., PROHASKA R.;  
RA BIOCHIM. BIOPHYS. ACTA 1151:149-152(1993).  
CC -1- FUNCTION: THOUGHT TO REGULATE CATION CONDUCTANCE.  
CC -1- SUBCELLULAR LOCATION: EXPOSED ON THE CYTOPLASMIC SURFACE OF THE  
CC MEMBRANE.  
CC -1- DISEASE: DEFECTS IN EPB72 IN RED CELLS OF PATIENTS WITH HEREDITARY  
CC SROMAOCYTOSIS OR CRYOHYOCYTOSIS RESULTS IN AN INCREASED NA+/K+-  
CC PERMEABILITY AND HENCE TO A DISORDER OF CELL VOLUME CONTROL.  
CC -1- SIMILARITY: BELONGS TO THE BAND 7/MCC-2 FAMILY.  
CC  
CC EMBL: X60067; G31069; -.  
DR EMBL: X85116; E140838; -.  
DR EMBL: X85117; E140838; JOINED.  
DR EMBL: M81635; G181184; -.  
DR EMBL: U33931; G1103842; -.  
DR EMBL: U33923; G1103842; JOINED.  
DR EMBL: U33926; G1103842; JOINED.  
DR EMBL: U33927; G1103842; JOINED.  
DR EMBL: U33928; G1103842; JOINED.  
DR EMBL: U33929; G1103842; JOINED.  
DR EMBL: U33930; G1103842; JOINED.  
DR PIR: S17659; S17659.  
DR MIM: 133090; -.  
DR MIM: 185000; -.  
DR PROSITE: PS01270; BAND\_7; 1.  
KW ERYTHROCYTE; TRANSMEMBRANE; PHOSPHORYLATION.  
FT INT\_MET 0  
FT TRANSMEM 25  
FT DOMAIN 54  
FT MOD\_RES 9  
FT CONFLICT 5  
FT SEQUENCE 287 AA; 31599 MM; 4789P0AD CRC32;  
Query Match 100.0%; Score 19; DB 1; Length 287?  
Best Local Similarity 25.0%; Pred. No. 2.88e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0.  
Db 199 PVDORAM 206  
QY 2 PXXXXXXM 9  
RESULT 4  
AC G3P\_ESCHE STANDARD; PRT; 294 AA.  
DT 01-MAR-1992 (REL. 21, CREATED)  
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)  
DE GLYCEALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH)  
DE (FRAGMENT).  
EN GAP.  
OS ESCHERICHIA HERMANNII.

Query Match 100.0%; Score 19; DB 1; Length 294;  
 Best Local Similarity 25.0%; Pred. No. 2.88e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 107 PSKDDPM 114  
 OY 2 PXXXXXXM 9

RESULT 5  
 AC P2164; STANDARD; PRT; 303 AA.  
 DT 01-MAR-1992 (REL. 21, CREATED)  
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)  
 DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH)  
 DE (FRAGMENT)  
 GN GAP.  
 OS KLEBSIELLA PNEUMONIAE.  
 OC PROKARYOTA: GRACILICUTES; SCOTOBACTERIA: FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 13883;  
 RX MEDLINE: 91319745.  
 RA NELSON K., WHITTAM T.S., SELANDER R.K.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:6667-6671(1991).  
 RN [2]  
 RP SEQUENCE OF 10-303 FROM N.A.  
 RC STRAIN-LD119;  
 RX MEDLINE: 92065252.  
 RA LAWRENCE J.G., OCHMAN H., HARTL D.L.;  
 RL J. GEN. MICROBIOL. 137:1911-1921(1991).  
 CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
 + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.  
 CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 DR EMBL: M68869; G149197; ALT\_TERM.  
 DR HSSP: P00354; 3GPD.  
 DR PROSITE: PS00071; GAPDH: 1.  
 KW GLYCOLYSIS; OXIDOREDUCTASE; NAD.  
 FT BINDING 144 144  
 FT ACT\_SITE 171 171  
 FT NON\_TER 303 303  
 SO SEQUENCE 303 AA; 32306 MW; 30FC255F CRC32;

Query Match 100.0%; Score 19; DB 1; Length 303;  
 Best Local Similarity 25.0%; Pred. No. 2.88e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 116 PSKDDPM 123  
 OY 2 PXXXXXXM 9

RESULT 6  
 ID BEL1\_SFV1 STANDARD; PRT; 308 AA.  
 AC P29169;  
 DT 01-DEC-1992 (REL. 24, CREATED)  
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)  
 DE BEL-1 OR TAF.  
 GN BEL-1 OR TAF.  
 OS SIMIAN FOAMY VIRUS (TYPE 1) (SFV-1).  
 OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 OC SPUNAVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE: 91276270.  
 RA KUPREC J.-J., KAY A., HAVAT M., RAVIER R., PERIES J., GALIBERT F.;  
 RL J. VIROL. 65:2903-2909(1991).  
 CC -1- FUNCTION: THIS PROTEIN IS A TRANSCRIPTIONAL TRANSACTIVATOR.  
 DR EMBL: X54482; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: M74039; G454845; -.  
 DR PIR: B39924; WMLJ51.  
 DR PIR: S18740; S18740.  
 KW TRANSCRIPTION REGULATION; ACTIVATOR.  
 FT CONFLICT 89 89  
 FT CONFLICT 119 119  
 FT CONFLICT 257 257  
 SO SEQUENCE 308 AA; 35311 MW; 96D2D7B2 CRC32;

Query Match 100.0%; Score 19; DB 1; Length 308;  
 Best Local Similarity 25.0%; Pred. No. 2.88e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 266 PEVGSPEM 273  
 OY 2 PXXXXXXM 9

RESULT 7  
 ID CCR4\_RAT STANDARD; PRT; 349 AA.  
 AC 008565;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE SDF-1 RECEPTOR (STROMAL CELL-DERIVED FACTOR 1 RECEPTOR) (FUSIN) (CXCR-4) (LEUCOCYTE-DERIVED SEVEN TRANSMEMBRANE DOMAIN RECEPTOR) (LESTR).  
 GN CXCR4 OR CXCR4.  
 OS RATIUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WISPAR; TISSUE-SPLEEN;  
 RA HARRISON J.K., SALAFRANCA M.N.;  
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: RECEPTOR FOR THE C-X-C CHEMOKINE SDF-1. TRANSDUCES A  
 SIGNAL BY INCREASING THE INTRACELLULAR CALCIUM IONS LEVEL.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 DR EMBL: U90610; G1906613; -.  
 DR PROSITE: PS00237; G\_PROTEIN\_RECEPTOR; 1.

KM G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN.  
 FT DOMAIN 1 36  
 FT TRANSMEM 37 60  
 FT TRANSMEM 61 76  
 FT TRANSMEM 77 96  
 FT TRANSMEM 97 107  
 FT TRANSMEM 108 129  
 FT TRANSMEM 130 151  
 FT TRANSMEM 152 172  
 FT TRANSMEM 173 197  
 FT TRANSMEM 198 217  
 FT TRANSMEM 218 237  
 FT TRANSMEM 238 258  
 FT TRANSMEM 259 282  
 FT TRANSMEM 283 302  
 FT TRANSMEM 303 349  
 FT DISULFID 106 183  
 FT CARBOHYD 8  
 SO SEQUENCE 349 AA; 39334 MW; CBA6532 CRC32;

Query Match  
 Best Local Similarity 100.0%; Score 19; DB 1; Length 349;  
 Best Local Similarity 25.0%; Pred. No. 2.88e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 89 PFWAVDAM 96  
 Oy 2 PXXXXXXM 9

RESULT 8  
 ID CCR4\_MOUSE STANDARD; PRT; 359 AA.  
 AC P70658; P70346; 009062; 009059;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DE SDF-1 RECEPTOR (STROMAL CELL-DERIVED FACTOR 1 RECEPTOR) (FUSIN) (CXCR-4) (LEUKOCYTE-DERIVED SEVEN TRANSMEMBRANE DOMAIN RECEPTOR) (LESTR).  
 GN CXCR4 OR LESTR OR CXCR4 OR SDF1R.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6 X CBA; TISSUE-THYMUS;  
 RA MOERPS B.; FRODL R.; KESSLER H.; GIESCHIK P.;  
 RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-129/SV;  
 RA HESEN M.; BERMAN M.A.; GERARD C.; DORF M.E.;  
 RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE; 97113334.  
 RA HESEN M.; BERMAN M.A.; BENSON J.D.; GERARD C.; DORF M.E.;  
 RL J. IMMUNOL. 157:5455-5460(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BONE MARROW;  
 RA NAKASAWA T.; NAKAJIMA T.; TACHIBANA K.; IIZASA H.; BLEUL C.C.;  
 RL YOSHIE O.; MATSUSHIMA K.; YOSHIDA N.; SPRINGER T.A.; KISHIMOTO T.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:14726-14729(1996).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6; TISSUE-THYMUS;  
 RA SUZUKI G.; NAKATA Y.; UZAWA A.; SHIRASAWA T.; SAITO T.; MITA K.;  
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: RECEPTOR FOR THE C-X-C CHEMOKINE SDF-1. TRANSDUCES A SIGNAL BY INCREASING THE INTRACELLULAR CALCIUM IONS LEVEL.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

DR EMBL; X99581; E281127; -;  
 DR EMBL; X99582; E281126; -;  
 DR EMBL; U59760; G1527135; -;  
 DR EMBL; U65580; G1731651; -;  
 DR EMBL; D87747; G1772446; -;  
 DR EMBL; AB000803; G1816446; -;  
 DR MGD; MG1:109563; CMKAR4.  
 DR PROSITE; PS00237; G-PROTEIN\_RECEPTOR; 1.  
 KM G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN.  
 FT DOMAIN 1 41  
 FT TRANSMEM 42 65  
 FT TRANSMEM 66 81  
 FT TRANSMEM 82 101  
 FT TRANSMEM 102 112  
 FT TRANSMEM 113 134  
 FT TRANSMEM 135 156  
 FT TRANSMEM 157 177  
 FT TRANSMEM 178 207  
 FT TRANSMEM 208 227  
 FT TRANSMEM 228 247  
 FT TRANSMEM 248 268  
 FT TRANSMEM 269 292  
 FT TRANSMEM 293 312  
 FT TRANSMEM 313 359  
 FT DISULFID 111 193  
 FT CARBOHYD 13  
 FT CONFLICT 216 216  
 SO SEQUENCE 359 AA; 40426 MW; 1037B4D3 CRC32;

Query Match  
 Best Local Similarity 100.0%; Score 19; DB 1; Length 359;  
 Best Local Similarity 25.0%; Pred. No. 2.88e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 94 PFWAVDAM 101  
 Oy 2 PXXXXXXM 9

RESULT 9  
 ID EDG2\_BOVIN STANDARD; PRT; 364 AA.  
 AC Q28031;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DE PROBABLE G PROTEIN-COUPLED RECEPTOR EDG-2 (REL. 3).  
 GN EDG2.  
 OS BOS TAURUS (BOVINE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RA MACRAE A.D.; FREMONT R.T.; JABER M.; PETERSEN A.S.; LEFKOWITZ R.J.;  
 RL BRAIN RES. MOL. BRAIN RES. 42:245-254(1996).  
 CC -1- FUNCTION: ORPHAN RECEPTOR.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 DR EMBL; U48236; G1203901; -;  
 DR PROSITE; PS00237; G-PROTEIN\_RECEPTOR; 1.  
 KM G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;  
 FT DOMAIN 1 50  
 FT TRANSMEM 51 75  
 FT TRANSMEM 76 82  
 FT TRANSMEM 83 111  
 FT TRANSMEM 112 125  
 FT TRANSMEM 126 144  
 FT TRANSMEM 145 163  
 FT TRANSMEM 164 189  
 FT TRANSMEM 190 205  
 FT TRANSMEM 206 226  
 FT TRANSMEM 227 258  
 FT DOMAIN 227 258  
 CYTOPLASMIC (POTENTIAL).



FT TRANSMEM 259 280 6 (POTENTIAL).  
 FT DOMAIN 281 294 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 295 315 7 (POTENTIAL).  
 FT DOMAIN 316 364 CYTOPLASMIC (POTENTIAL).  
 FT LIPID 327 327 PALMITATE (BY SIMILARITY).  
 FT CARBOHYD 27 27 POTENTIAL.  
 FT CARBOHYD 35 35 POTENTIAL.  
 SQ SEQUENCE 364 AA; 41070 MW; A237BB94 CRC32;

Query Match 100.0%; Score 19; DB 1; Length 364;  
 Best Local Similarity 25.0%; Pred. No. 2.88e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 246 PRNRDPTM 253  
 QY 2 PXXXXXXM 9

RESULT 10  
 ID ACT2.OXYNO STANDARD; PRT: 375 AA.  
 AC P55805;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ACTIN, CYTOPLASMIC (ACTIN, MICRONUCLEAR).  
 GN MIC-ACT-1 AND MIC-ACT-2.  
 OS OXYTRICHA NOVA.  
 CC OXYTRICHA NOVA.  
 CC EUKARYOTA; PROTOZOA; CILIOPHORA; CILIATA; SPIROTRICHA; HYPOTRICHIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89345644.  
 RA GRESLIN A.F., PRESOTT D.M., OKA Y., LOUKIN S.H., CHAPPELL J.C.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 86:6264-6268(1989).  
 CC -1- FUNCTION: ACTINS ARE HIGHLY CONSERVED PROTEINS THAT ARE INVOLVED  
 IN VARIOUS TYPES OF CELL MOTILITY AND ARE UBQUITOUSLY EXPRESSED  
 IN ALL EUKARYOTIC CELLS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 DR EMBL; M25531; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; M25530; -; NOT\_ANNOTATED\_CDS.  
 DR PROSITE; PS00406; ACTINS\_1; 1.  
 DR PROSITE; PS00432; ACTINS\_2; 1.  
 DR PROSITE; PS01132; ACTINS\_ACT\_LIKE; 1.  
 KW STRUCTURAL PROTEIN; MULTIGENE FAMILY.  
 SO SEQUENCE 375 AA; 41967 MW; 00CC922F CRC32;

Query Match 100.0%; Score 19; DB 1; Length 375;  
 Best Local Similarity 25.0%; Pred. No. 2.88e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 111 PKNRKM 118  
 QY 2 PXXXXXXM 9

RESULT 11  
 ID ACT2.OXYTR STANDARD; PRT: 375 AA.  
 AC P53469;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ACTIN, CYTOPLASMIC (ACTIN, MICRONUCLEAR).  
 OS OXYTRICHA TRIFALIA.  
 CC EUKARYOTA; PROTOZOA; CILIOPHORA; CILIATA; SPIROTRICHA; HYPOTRICHIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-WR;  
 RX MEDLINE; 95249578.  
 RA DUBOIS M., PRESOTT D.M.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 92:3888-3892(1995).  
 CC -1- FUNCTION: ACTINS ARE HIGHLY CONSERVED PROTEINS THAT ARE INVOLVED  
 IN VARIOUS TYPES OF CELL MOTILITY AND ARE UBQUITOUSLY EXPRESSED  
 IN ALL EUKARYOTIC CELLS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

DR EMBL; U19288; G632547; -;  
 DR PROSITE; PS00406; ACTINS\_1; 1.  
 DR PROSITE; PS00432; ACTINS\_2; 1.  
 DR PROSITE; PS01132; ACTINS\_ACT\_LIKE; 1.  
 KW STRUCTURAL PROTEIN.  
 SQ SEQUENCE 375 AA; 41966 MW; 89EC22A1 CRC32;

Query Match 100.0%; Score 19; DB 1; Length 375;  
 Best Local Similarity 25.0%; Pred. No. 2.88e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 111 PKNRKM 118  
 QY 2 PXXXXXXM 9

RESULT 12  
 ID ACT1.SOLTV STANDARD; PRT: 377 AA.  
 AC P30167;  
 DT 01-APR-1993 (REL. 25, CREATED)  
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE ACTIN 58.  
 GN ACS8.  
 OS SOLANUM TUBEROSUM (POTATO).  
 CC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;  
 CC SOLANALES; SOLANACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. MARIS PIPER; TISSUE-LEAF.  
 RX MEDLINE; 91012599.  
 RA DROUIN G., DOVER G.A.;  
 RL J. MOL. EVOL. 31:132-150(1990).  
 CC -1- FUNCTION: ACTINS ARE HIGHLY CONSERVED PROTEINS THAT ARE INVOLVED  
 IN VARIOUS TYPES OF CELL MOTILITY AND ARE UBQUITOUSLY EXPRESSED  
 IN ALL EUKARYOTIC CELLS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 DR EMBL; X55749; G21536; -;  
 DR PIR; S20094; S20094.  
 DR HSP; P02570; 2BT.  
 DR PROSITE; PS00406; ACTINS\_1; 1.  
 DR PROSITE; PS00432; ACTINS\_2; 1.  
 DR PROSITE; PS01132; ACTINS\_ACT\_LIKE; 1.  
 KW STRUCTURAL PROTEIN; MULTIGENE FAMILY.  
 SO SEQUENCE 377 AA; 41786 MW; F986FEDD CRC32;

Query Match 100.0%; Score 19; DB 1; Length 377;  
 Best Local Similarity 25.0%; Pred. No. 2.88e+02;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 114 PKNRKM 121  
 QY 2 PXXXXXXM 9

RESULT 13  
 ID ACT1.SORVU STANDARD; PRT: 377 AA.  
 AC P53504;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE ACTIN 1.  
 GN AC1.  
 OS SORGHUM VULGARE (SORGHUM).  
 CC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;  
 CC CYPERALES; GRAMINEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. JINZHONG; TISSUE-LEAF;  
 RA ZHOU L.;  
 RL SUBMITTED (JUN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACTINS ARE HIGHLY CONSERVED PROTEINS THAT ARE INVOLVED  
 IN VARIOUS TYPES OF CELL MOTILITY AND ARE UBQUITOUSLY EXPRESSED

```

CC      IN ALL EUKARYOTIC CELLS.
CC      -1- FUNCTION: ESSENTIAL COMPONENT OF CELL CYTOSKELETON; PLAYS AN
CC      IMPORTANT ROLE IN CYTOPLASMIC STREAMING, CELL SHAPE DETERMINATION,
CC      CELL DIVISION, ORANELLE MOVEMENT AND EXTENSION GROWTH.
CC      -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
DR      EMBL: X79378; G499012; -.
DR      PROSITE: PS00406; ACTINS_1; 1.
DR      PROSITE: PS00432; ACTINS_2; 1.
DR      PROSITE: PS01132; ACTINS_ACT LIKE; 1.
KW      STRUCTURAL PROTEIN; MULTIGENE FAMILY.
SQ      SEQUENCE 377 AA; 41863 MW; 63C30BA1 CRC32;

Query Match
Best Local Similarity 25.0%; Score 19; DB 1; Length 377;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db      114 PKANREXM 121
QY      2 PXXXXXXM 9

RESULT 14
ID      EMBL_ADENT      STANDARD:      PRT:      391 AA.
AC      P04885;
DT      13-AUG-1987 (REL. 05, CREATED)
DT      13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT      01-JUL-1989 (REL. 11, LAST ANNOTATION UPDATE)
DE      EARLY E1B 44 KD PROTEIN.
OS      TUPAIA ADENOVIRUS.
OC      VIRIDAE; DS-DNA NONENVELOPED VIRUSES; ADENOVIRIDAE; MASTADENOVIRUSES.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE: 85232053.
RA      FLUGEL R.M., BANNERT H., SUHAI S., DARAI G.;
RL      GENE 34:73-80(1985).
DR      EMBL: M10054; G210027; -.
DR      PIR: A03813; ERADT4.
KM      EARLY PROTEIN.
SQ      SEQUENCE 391 AA; 43541 MW; 18A52348 CRC32;

Query Match
Best Local Similarity 25.0%; Score 19; DB 1; Length 391;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db      57 PDDDWITDM 64
QY      2 PXXXXXXM 9

RESULT 15
ID      EFTU_CYCME      STANDARD:      PRT:      409 AA.
AC      P50373;
DT      01-OCT-1996 (REL. 34, CREATED)
DT      01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT      01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE      ELONGATION FACTOR TU (EF-TU).
GN      TUF.
OS      CYCLOTHELLA MENECHINIANA.
OC      CHLOROPLAST.
OC      EUKARYOTA; PLANTA; PHYCOPHYTA; BACILLARIOPHYTA (DIATOMS).
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE: 95392822.
RA      DELWICHE C.F., KUHSEL M., PALMER J.D.;
RL      MOL. PHYLOGENET. EVOL. 4:110-128(1995).
CC      -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC      AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC      BIOSYNTHESIS.
CC      -1- SUBCELLULAR LOCATION: CHLOROPLAST.
DR      EMBL: U09430; G836836; -.
DR      PROSITE: PS00301; EFACOR_GTP; 1.
KW      ELONGATION FACTOR; PROTEIN BIOSYNTHESIS; CHLOROPLAST;
KW      GTP-BINDING.

```

```

FT      NP_BIND      19      26      GTP (BY SIMILARITY).
FT      NP_BIND      81      85      GTP (BY SIMILARITY).
FT      NP_BIND      136     139      GTP (BY SIMILARITY).
SQ      SEQUENCE 409 AA; 44605 MW; 980AF3E5 CRC32;

Query Match
Best Local Similarity 25.0%; Score 19; DB 1; Length 409;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db      381 PVAIEBGM 388
QY      2 PXXXXXXM 9

Search completed: Fri Sep 11 14:00:51 1998
Job time : 6 secs.

```

(TM)

Distribution rights by Oxford Molecular Ltd

in - protein database sec

104.702 Million cell updates/sec

100

(1-9) from US08452843.pep

6 XXXXXM 9

PAM 150

140555 seqs, 42109429 residues

Malchuk

1:sn final 2:sn human 3:sn invertabrate 4:sn mammal

9:sp bacteria 10:sp rodent 11:sp virus 12:sp vertebrate

1

Mean 11.553; Variance 3.929; scale 2.941

ved by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB	ID	Description	Pred. No.
1	19	100.0	42	015948	THYOGLOBULIN (PRAGEN	7.49e+01
2	19	100.0	106	3 P91139	COSMID C37H5.	7.49e+01
3	19	100.0	170	3 012952	COSMID F05E10.	7.49e+01
4	19	100.0	171	3 022735	SIMILAR TO C. TENTANS	7.49e+01
5	19	100.0	175	3 022954	COSMID F10G2.	7.49e+01
6	19	100.0	226	3 020421	F45D3.3.	7.49e+01
7	19	100.0	276	3. 017808	COSMID C08A9.	7.49e+01
8	19	100.0	304	3 023665	ZK930.2.	7.49e+01
9	19	100.0	317	3 019628	COSMID F20B6.	7.49e+01
10	19	100.0	326	3 094603	SIMILAR TO MITOCHONDRI	7.49e+01
11	19	100.0	334	3 018353	C32c4.3.	7.49e+01
12	19	100.0	378	3 018852	HYPOTHETICAL PROTEIN C	7.49e+01
13	19	100.0	355	3 018659	SIMILAR TO THYROTRONIN	7.49e+01
14	19	100.0	400	3 022904	SIMILARITY TO M. GENIT	7.49e+01
15	19	100.0	414	3 020841	SIMILAR TO RHODOPIN	7.49e+01
16	19	100.0	429	1 013326	HOMOCYSTEINE SYNTHASE	7.49e+01
17	19	100.0	429	3 018682	SIMILAR TO CASEIN KINA	7.49e+01
18	19	100.0	437	3 021450	MO1F1.1.	7.49e+01
19	19	100.0	478	3 007408	GLUCALPHA2B PROTEIN P	7.49e+01
20	19	100.0	430	2 005409	CHECKPOINT SUPPRESSOR	7.49e+01

21	19	100.0	494	3	P15155	COSMID T82A11..	7.49e+01
22	19	100.0	494	3	Q24737	ALPHA-AMYLASE.	7.49e+01
23	19	100.0	500	3	I18475	CS3A5..7.	7.49e+01
24	19	100.0	510	3	Q20632	CODED FOR BY C. ELEGAN	7.49e+01
25	19	100.0	525	3	Q22462	TJ3H10.1 (FRAGMENT).	7.49e+01
26	19	100.0	554	3	I19253	SIMILAR TO TYROSINE-PR	7.49e+01
27	19	100.0	599	3	Q24068	REF(2) P PROTEIN.	7.49e+01
28	19	100.0	599	3	Q24069	REF(2) P PROTEIN.	7.49e+01
29	19	100.0	640	2	Q02297	HREGULIN-ALPHA (HRG-A	7.49e+01
30	19	100.0	657	3	I17547	GLUCALPHA2A PROTEIN.	7.49e+01
31	19	100.0	662	3	Q23318	2C34.9.	7.49e+01
32	19	100.0	666	3	P90858	FE8G11.3.	7.49e+01
33	19	100.0	676	2	O15296	1S-LIPOXYGENASE.	7.49e+01
34	19	100.0	684	3	P91132	CODED FOR BY C. ELEGAN	7.49e+01
35	19	100.0	720	3	Q25258	LDI ELEMENT ORC, D, A	7.49e+01
36	19	100.0	726	3	Q27936	AMP-BINDING CASSETTE P	7.49e+01
37	19	100.0	862	3	Q22354	SIMILAR TO HUMAN VAV G	7.49e+01
38	19	100.0	916	3	Q22276	HYPOTHEETICAL PROTEIN T	7.49e+01
39	19	100.0	926	3	Q19853	CODED FOR BY C. ELEGAN	7.49e+01
40	19	100.0	1021	3	Q20217	FO0F1.5.	7.49e+01
41	19	100.0	1121	2	Q16821	SERINE /THROMBIN SPEC	7.49e+01
42	19	100.0	1303	3	Q22332	OSB11.2 (FRAGMENT)	7.49e+01
43	19	100.0	1793	1	O13555	YHR214C-BP.	7.49e+01
44	19	100.0	1862	3	Q20090	C. ELEGANS DNA-DIRECTE	7.49e+01
45	19	100.0	2447	3	Q22463	SIMILAR TO ZINC FINGER	7.49e+01

## ALIGNMENTS

```

RESULT      1
ID           015948
AC           015948:
DT           01-NOV-1996 (TREMBL:REL. 01, CREATED)
DT           01-NOV-1996 (TREMBL:REL. 01, LAST SEQUENCE UPDATE)
DT           01-NOV-1996 (TREMBL:REL. 01, LAST ANNOTATION UPDATE)
DE           THYROGLOBULIN (FRAGMENT).
OS           HOMO SAPIENS (HUMAN).
OC           EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:
OC           EUHERIA: PRIMATES.
RN           [1]
RP           SEQUENCE FROM N.A.
RA           TAGOVNIK H.M., COCHAUX P., CORACH D., VASSART G.;
RL           MOL. CELL. ENDOCRINOL. 84:1-6(1992).
DR           EMBL; S4807; G252171; -.
FT           NON TER
SQ           SEQUENCE 42 AA; 4734 MW; 58DDBA63 CRC32;

Query Match      100.0%; Score 19; DB 2; Length 42;
Best Local Similarity 25.0%; Pred. NO. 7.49e+01;
Matches          2; Conservative          0; Mismatches 6; Indels 0; Gaps 0

QY      34 PDSEFPVM 41
        |
        2 PXXXXXXM 9

RESULT      2
ID           P91139
AC           P91139:
DT           01-MAY-1997 (TREMBL:REL. 03, CREATED)
DT           01-MAY-1997 (TREMBL:REL. 03, LAST SEQUENCE UPDATE)
DT           01-MAY-1997 (TREMBL:REL. 03, LAST ANNOTATION UPDATE)
DE           COSMID C37H5.
GN           C37H5.11.
GN           CANONHABDITIS ELEGANS.
OC           EUKARYOTA: METAZOA: ACCOELOMATES: NEMATODA: SECERNENTEA: RHADITIDA.
RN           [1]
RP           SEQUENCE FROM N.A.
RC           STRAIN-BRISTOL N2;
RX           MEDLINE; 94150718.
RA           WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKKS M.,
RA           BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA           CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON I.,

```

```

ID      4      PRELIMINARY;  PRT;  171 AA.
RD      022739
AC      022739;
DT      01-NOV-1996 (TREMBLREL. 01, CREATED)
DT      01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT      01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE      SIMILAR TO C. TENTANS SECRETORY PROTEIN 2.
GN      T24D8.5.
OS      CAENORHABDITIS ELEGANS.
OC      EMBAROTLA. METAODA. ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 94150718.
RA      WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA      BOFIELD J., BURTON M., CONNELL M., COPSEY T., COOPER J.,
RA      COULSON A., CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,
RA      FUTON L., GARDNER A., GREEN P., HAKINS T., HILLIER L., JIER M.,
RA      JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
RA      LARREILLE P., LIGHNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,
RA      O'CALLAGHAN M., PARSONS J., PERCY C., RIFKEN L., ROOPRA A.,
RA      SANDERS D., SHOWNKEEN R., SVALDON N., SMITH A., SONNHAMMER E.,
RA      STADEN R., SULSTON J., THERRY-MIEG J., THOMAS K., VAUDIN M.,
RA      VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA      WILKINSON-SPROAT J., WOHLDMAN P.;
RL      NATURE 368:32-38(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      MARTIN J.;
RL      SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR      EMBL; 040422; G1065532; -
SD      SEQUENCE 171 AA; 19405 MW; 3530FA69 CAC32;

Query Match      100.0%; Score 19; DB 3; Length 171;
Best Local Similarity 25.0%; Pred. No. 7.49e+01;
Matches      2; Conservative      0; Mismatches      6; Indels      0; Gaps      0;

Db      150 PGKRSNDK 157
Qy      2 PXXXXXX 9

RESULT      5
ID      022954      PRELIMINARY;  PRT;  175 AA.
AC      022954;
DT      01-NOV-1996 (TREMBLREL. 01, CREATED)
DT      01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT      01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE      COSMID F10G2.
DE      F10G2.4.
OS      CAENORHABDITIS ELEGANS.
OC      EMBAROTLA. METAODA. ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 94150718.
RA      WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA      BOFIELD J., BURTON M., CONNELL M., COPSEY T., COOPER J.,
RA      COULSON A., CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,
RA      FUTON L., GARDNER A., GREEN P., HAKINS T., HILLIER L., JIER M.,
RA      JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
RA      LARREILLE P., LIGHNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,
RA      O'CALLAGHAN M., PARSONS J., PERCY C., RIFKEN L., ROOPRA A.,
RA      SANDERS D., SHOWNKEEN R., SVALDON N., SMITH A., SONNHAMMER E.,
RA      STADEN R., SULSTON J., THERRY-MIEG J., THOMAS K., VAUDIN M.,
RA      VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA      WILKINSON-SPROAT J., WOHLDMAN P.;
RL      NATURE 368:32-38(1994).

```

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA MURRAY J., WOHLDMAN P.;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U64836; G1458256;  
SQ SEQUENCE 175 AA; 19709 MW; 78F478D0 CRC32;

Query Match  
Best Local Similarity 25.0%; Score 19; DB 3; Length 175;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 46 PASSSPDM 53  
QY 2 PXXXXXXM 9

RESULT 6  
ID Q20421 PRELIMINARY; PRT: 236 AA.  
AC Q20421;  
DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMELREL. 01, LAST ANNOTATION UPDATE)  
DE F453.3.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECCERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA WHITE S.;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,  
RA COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,  
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,  
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFKEN L., ROOPRA A.,  
RA SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONNHAMMER E.,  
RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 368:32-38(1994).  
DR EMBL: 278063; E257910;  
SQ SEQUENCE 236 AA; 25069 MW; 6AB5CE27 CRC32;

Query Match  
Best Local Similarity 100.0%; Score 19; DB 3; Length 236;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 85 PFTGGM 92  
QY 2 PXXXXXXM 9

RESULT 7  
ID Q17808 PRELIMINARY; PRT: 276 AA.  
AC Q17808;  
DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMELREL. 01, LAST ANNOTATION UPDATE)  
DE COSMID C08A9.  
GN C08A9.5.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECCERNENTEA; RHABDITIDA.  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,  
RA COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,  
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,  
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFKEN L., ROOPRA A.,  
RA SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONNHAMMER E.,  
RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA LATREILLE P.;  
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA WATERSTON R.;  
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U42844; G1125801;  
SQ SEQUENCE 276 AA; 30926 MW; B65B60BD CRC32;

Query Match  
Best Local Similarity 100.0%; Score 19; DB 3; Length 276;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 156 PSEPRYM 163  
QY 2 PXXXXXXM 9

RESULT 8  
ID Q23665 PRELIMINARY; PRT: 304 AA.  
AC Q23665;  
DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMELREL. 01, LAST ANNOTATION UPDATE)  
DE ZK930.2.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECCERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SWINBURNE J.;  
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,  
RA COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,  
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,  
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFKEN L., ROOPRA A.,  
RA SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONNHAMMER E.,  
RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 368:32-38(1994).  
DR EMBL: 270213; E229232;  
SQ SEQUENCE 304 AA; 34533 MW; C84A715D CRC32;

Query Match  
Best Local Similarity 100.0%; Score 19; DB 3; Length 304;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 273 PTCPCWK 280  
QY 2 PXXXXXXM 9

RESULT 9  
ID Q19628 PRELIMINARY: PRT: 317 AA.  
AC Q19628;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE COSMID F20B6.  
GN F20B6.5.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,  
RA COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,  
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,  
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIKKEN L., ROOPRA A.,  
RA SANDERS D., SHOWKNEEN R., SMALDON N., SMITH A., SONNHAMMER E.,  
RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.,  
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MIX P.;  
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U41015; G1086647;  
SQ SEQUENCE 317 AA; 35814 MW; 22047722 CRC32;

Query Match 100.0%; Score 19; DB 3; Length 317;  
Best Local Similarity 25.0%; Pred. No. 7.49e+01;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 56 PPGCRSM 63  
QY 2 PXXXXXX 9

RESULT 10  
ID Q94603 PRELIMINARY: PRT: 326 AA.  
AC Q94603;  
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)  
DE SIMILAR TO MITOCHONDRIAL TRICARBOXYLATE CARRIER.  
OS LEISHMANIA MAJOR.  
OC EUKARYOTA; PROTISTA; SARCOMASTIGOPHORA; MASTIGOPHORA; KINETOPLASTIDA;  
RN TRYPANOSOMATIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MYLER P.J., STUART K.D., MAGNESS C., DEVOS T., WESTLAKE T.,  
RA LEMLEY C., RILEY P.,  
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U70253; G1617564;  
SQ SEQUENCE 326 AA; 36141 MW; A629D498 CRC32;

Query Match 100.0%; Score 19; DB 3; Length 326;  
Best Local Similarity 25.0%; Pred. No. 7.49e+01;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 84 PPGCRSM 91

QY 2 PXXXXXX 9

RESULT 11  
ID Q18353 PRELIMINARY: PRT: 334 AA.  
AC Q18353;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE C32C4.3.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MCMURRAY A.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,  
RA COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,  
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,  
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIKKEN L., ROOPRA A.,  
RA SANDERS D., SHOWKNEEN R., SMALDON N., SMITH A., SONNHAMMER E.,  
RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.,  
RL NATURE 368:32-38(1994).  
DR EMBL: Z73905; E249682;  
SQ SEQUENCE 334 AA; 36101 MW; 671589E7 CRC32;

Query Match 100.0%; Score 19; DB 3; Length 334;  
Best Local Similarity 25.0%; Pred. No. 7.49e+01;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 80 PUYETFM 87  
QY 2 PXXXXXX 9

RESULT 12  
ID Q18852 PRELIMINARY: PRT: 378 AA.  
AC Q18852;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL PROTEIN C54G4.7.  
GN C54G4.7.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,  
RA COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,  
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,  
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIKKEN L., ROOPRA A.,  
RA SANDERS D., SHOWKNEEN R., SMALDON N., SMITH A., SONNHAMMER E.,  
RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.,  
RL NATURE 368:32-38(1994).  
DR EMBL: Z75533; E348482;  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 378 AA; 43184 MW; D8B545A3 CRC32;

Query Match 100.0%; Score 19; DB 3; Length 378;

RA BONEFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COLLISON A.  
RA CRAXTON M., DEAR S., DU 2., DURBIN R., FAVELLO A., FULTON L.,

Query Match	100.0%; Score 19; DB 3; Length 414;
-------------	-------------------------------------

Query Match	100.0%;	Score 19;	DB 3;	Length 414;
Best Local Similarity	25.0%;	Pred. No. 7.49e+01;		
Matches	2;	Conservative	0;	Mismatches 6; Indels 0; Gaps 0;

Sun Sep 13 10:56:47 1998

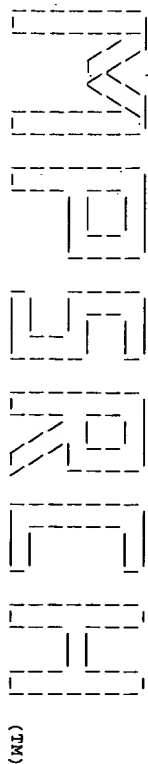
US-08-452-843-30.rpt

Page 6

Db 78 PYTKKATM 85  
Oy 2 PXXXXXXM 9

Search completed: Fri Sep 11 14:01:48 1998  
Job time : 40 secs.





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:55:53 1998; MasPar time 2.57 Seconds

Tabular output not generated. 56.700 Million cell updates/sec

Title: >US-08-452-843-21  
Description: (1-9) from US08452843.pcp  
Perfect Score: 58  
Sequence: 1 GSRHSHSL 9

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match: 08  
Listing first 45 summaries

Database:

a-geneseg32  
1:part1 2:part2 3:part3 4:part4 5:parts 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 14.620; Variance 35.727; scale 0.409

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	58	100.0	9	18	R89382	p53 derived immunogen	4.96e-01
2	58	100.0	15	10	R54917	Immunodominant epitope	4.96e-01
3	58	100.0	15	10	R54916	Immunodominant epitope	4.96e-01
4	58	100.0	26	21	W05367	Peptide p53PC360-386D	4.96e-01
5	58	100.0	27	21	W05367	Peptide p53PC360-386D	4.96e-01
6	58	100.0	27	21	W05367	Residues 360-386 of p	4.96e-01
7	58	100.0	27	21	W05367	Peptide p53PC360-386D	4.96e-01
8	58	100.0	28	21	W05365	Peptide p53PC360-386	4.96e-01
9	58	100.0	74	21	W09322	C-terminal domain of	4.96e-01
10	58	100.0	157	10	R51878	Human p53 amino acids	4.96e-01
11	58	100.0	328	10	R51876	Human p53 amino acids	4.96e-01
12	58	100.0	353	24	W28493	Human p53 protein var	4.96e-01
13	58	100.0	353	24	W28494	Human p53 protein var	4.96e-01
14	58	100.0	354	10	R51874	Human p53 amino acids	4.96e-01
15	58	100.0	393	19	W02617	Human p53 tumour supp	4.96e-01
16	58	100.0	393	18	R91933	Wild type p53 protein	4.96e-01
17	58	100.0	393	21	W13970	Modified p53 variant	4.96e-01
18	58	100.0	393	21	W05349	Human p53 mutant R273	4.96e-01

19	58	100.0	393	21	W13969	Modified p53 variant	4.96e-01
20	58	100.0	393	16	R94623	Modified p53 variant	4.96e-01
21	58	100.0	393	5	R22238	p53 protein.	4.96e-01
22	58	100.0	393	5	R22238	Sequence of 53 kd cel	4.96e-01
23	58	100.0	393	21	W05344	p53.	4.96e-01
24	58	100.0	393	21	W05345	Human p53.	4.96e-01
25	58	100.0	393	22	W13951	Human p53 mutant N239	4.96e-01
26	58	100.0	393	22	W13951	Human tumour-derived	4.96e-01
27	58	100.0	393	22	W13980	Human tumour-derived	4.96e-01
28	58	100.0	393	22	W13978	Human tumour-derived	4.96e-01
29	58	100.0	393	22	W13948	Human wild-type p53 t	4.96e-01
30	58	100.0	393	22	W13949	T284R modified human	4.96e-01
31	58	100.0	393	22	W13953	T284K modified human	4.96e-01
32	58	100.0	393	22	W13979	Human tumour-derived	4.96e-01
33	58	100.0	393	14	R79658	Human p53 protein.	4.96e-01
34	58	100.0	393	22	W13981	Human tumour-derived	4.96e-01
35	58	100.0	393	22	W13952	Human tumour-derived	4.96e-01
36	58	100.0	401	24	W28487	Human p53 protein var	4.96e-01
37	58	100.0	401	24	W28488	Human p53 protein var	4.96e-01
38	58	100.0	402	21	W13965	Chimeric p53 protein.	4.96e-01
39	58	100.0	404	21	W13963	Chimeric p53 protein.	4.96e-01
40	58	100.0	406	21	W13966	Chimeric p53 protein.	4.96e-01
41	58	100.0	406	21	W13964	Chimeric p53 protein.	4.96e-01
42	58	100.0	411	21	W13967	Chimeric p53 protein.	4.96e-01
43	58	100.0	438	10	R50088	p53 tumour suppressor	4.96e-01
44	58	100.0	438	14	R74272	Tumour suppressor pro	4.96e-01
45	58	100.0	533	23	W19763	p53-GM-CSF immunostim	4.96e-01

## ALIGNMENTS

RESULT 1  
ID R89382 standard; peptide: 9 AA.  
AC R89382;  
DE 18-SEP-1996 (first entry)  
DR p53 derived immunogenic peptide, residues 361-369.  
KW Therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN W05603140-A1.  
PD 08-FEB-1996.  
PE 21-JUL-1995; U09234  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI: 96-116784/12.  
PT Composn. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89382-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g. the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA:

Query Match 100.0%; Score 58; DB 18; Length 9;  
Best local similarity 100.0%; Pred. No. 4.96e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 gsrhshsl 9  
Oy 1 GSRHSHSL 9

RESULT 2

ID RS4917 standard; peptide; 15 AA.

AC RS4917;

DT 29-NOV-1994 (first entry)

DE Immunodominant epitope from p53 C-terminal.

KW cancer; pre-cancerous state; detection; diagnosis; human p53 gene;

KW immunodominant epitope; human cellular tumour antigen;

KW transformation-associated protein.

OS Homo sapiens.

PN WO9410306-A.

PD 11-MAY-1994.

PF 02-NOV-1998; F01082.

PR 02-NOV-1992; FR-013110.

PA (EURO-) LAB EUROBIOL SA.

PI Legros Y, Lubin R, Soussi T;

DR WPI; 94-167463/20.

PT New immuno:dominant epitope(s) of protein p53 - for detecting and monitoring antibodies indicative of cancer and precancerous

PS Claim 7; Page 43; 62pp; French.

CC Peptides derived from the N-terminal (amino acids 1-112) or the C-terminal (amino acids 350-393) of protein p53 which specifically react with anti-p53 antibodies in patients with cancer or

CC precancerous conditions are claimed. The peptides (RS4907-R54921) are useful for detecting and monitoring cancerous and precancerous conditions.

CC Sequence 15 AA:

Query Match 100.0%; Score 58; DB 10; Length 15;

Best Local Similarity 100.0%; Pred. No. 4.96e-01;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 6 gsrhsshl 14

QX 1 GSRHSSHL 9

RESULT 3

ID RS4916 standard; peptide; 15 AA.

AC RS4916;

DT 29-NOV-1994 (first entry)

DE Immunodominant epitope from p53 C-terminal.

KW cancer; pre-cancerous state; detection; diagnosis; human p53 gene;

KW immunodominant epitope; human cellular tumour antigen;

KW transformation-associated protein.

OS Homo sapiens.

PN WO9410306-A.

PD 11-MAY-1994.

PF 02-NOV-1993; F01082.

PR 02-NOV-1992; FR-013110.

PA (EURO-) LAB EUROBIOL SA.

PI Legros Y, Lubin R, Soussi T;

DR WPI; 94-167463/20.

PT New immuno:dominant epitope(s) of protein p53 - for detecting and monitoring antibodies indicative of cancer and precancerous

PS Claim 7; Page 43; 62pp; French.

CC Peptides derived from the N-terminal (amino acids 1-112) or the C-terminal (amino acids 350-393) of protein p53 which specifically react with anti-p53 antibodies in patients with cancer or

CC precancerous conditions are claimed. The peptides (RS4907-R54921) are useful for detecting and monitoring cancerous and precancerous conditions.

CC Sequence 15 AA:

Query Match 100.0%; Score 58; DB 10; Length 15;

Best Local Similarity 100.0%; Pred. No. 4.96e-01;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 6 gsrhsshl 14

QX 1 GSRHSSHL 9

RESULT 4

ID W05368 standard; peptide; 26 AA.

AC W05368;

DT 30-APR-1997 (first entry)

DE Peptide p53pc360-386DF.

KW Human; p53; cell proliferation; cell death; regulator; tumour; psoriasis;

KW negative regulatory region; DNA damaging agent; transplant rejection;

KW abnormal cell proliferation; atherosclerosis; cancer; autoimmune disease;

KW arterial restenosis; immune response; apoptosis; inducer; therapy;

KW proliferating lymphocytes.

OS Synthetic.

PN WO9625434-A1.

PD 22-AUG-1996.

PF 16-FEB-1996; U01535.

PR 16-FEB-1995; US-392542.

PA (PARB ) BAYER CORP.

PI (WIST-) WISTAR INST.

PI Halazonetis T, Hartwig W;

DR WPI; 96-393345/39.

PT New human p53-isomorphous peptide(s) and peptide-mimetic cpds. - used for activating p53 function, e.g. for treating tumours, cancers,

PS psoriasis, etc

CC Disclosure: Page 12; 55pp; English.

CC W05365-W05374 represent examples of the p53 (see W05344 for full length wild type sequence) peptides of the invention. These sequences all have additions or deletions of residues from the wild type peptide fragments of the invention (see W05350-W05364). The p53 protein functions to regulate cell proliferation and cell death, and is mutated in more than half of all human tumours. These sequences are used to activate the DNA binding activity of wild type p53, and p53 mutants (see W05345-W05349).

CC The peptides of the invention consist of at least four sequential amino acids from a negative regulatory region which maps to residues 361-383 of p53. These sequences preferably contain four amino acids from a non-human p53 sequence, contain D-form amino acids, and can also be cyclic

CC peptides. The sequences retain the structural characteristics of the original peptides, but the modifications render them less susceptible to cleavage by proteases and exopeptidases. As these sequences activate p53 DNA binding, they can be used to identify p53 mutants. The peptides can also be used for treating a patient with a tumour expressing a p53 mutant whose ability to bind DNA may be activated by one of the peptides. They can also be used for treating conditions such as exposure to DNA damaging agents, abnormal cell proliferation characteristic of psoriasis, atherosclerosis, cancer, arterial restenosis, autoimmune diseases and undesirable immune responses accompanying rejection of a transplant. The peptides can also induce apoptosis of specific cells, such as proliferating lymphocytes.

CC Sequence 26 AA:

Query Match 100.0%; Score 58; DB 21; Length 26;

Best Local Similarity 100.0%; Pred. No. 4.96e-01;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 2 gsrhsshl 10

QX 1 GSRHSSHL 9

RESULT 5

ID W05367 standard; peptide; 27 AA.

AC W05367;

DT 30-APR-1997 (first entry)

DE Peptide p53pc360-386DF.

KW Human; p53; cell proliferation; cell death; regulator; tumour; psoriasis;

KW negative regulatory region; DNA damaging agent; transplant rejection;

KW abnormal cell proliferation; atherosclerosis; cancer; autoimmune disease;

KW arterial restenosis; immune response; apoptosis; inducer; therapy;

KW proliferating lymphocytes.

OS Synthetic.

PN WO9625434-A1.

PD 22-AUG-1996.

PF 16-FEB-1996; U01535.

PR 16-FEB-1995: US-392542.  
PA (FARB ) BAYER CORP.  
PA (WIST-) WISTAR INST.  
PI Halazonetis T, Hartwig W;  
DR WPI: 96-393345/39.  
PT New human p53-isomorph peptide(s) and peptide:mimetic cpds. - used  
PT for activating p53 function, e.g. for treating tumours, cancers,  
PT psoriasis, etc  
PS Disclosure: Page 12: 55pp: English.  
CC W05365-W05374 represent examples of the p53 (see W05344 for full length  
CC wild type sequence) peptides of the invention. These sequences all have  
CC additions or deletions of residues from the wild type peptide fragments  
CC of the invention (see W05350-W05364). The p53 protein functions to  
CC regulate cell proliferation and cell death, and is mutated in more than  
CC half of all human tumours. These sequences are used to activate the DNA  
CC binding activity of wild type p53, and p53 mutants (see W05345-W05349).  
CC The peptides of the invention consist of at least four sequential amino  
CC acids from a negative regulatory region which maps to residues 361-383 of  
CC p53. These sequences preferably contain four amino acids from a non-human  
CC p53 sequence, contain D-form amino acids, and can also be cyclic  
CC peptides. The sequences retain the structural characteristics of the  
CC original peptides, but the modifications render them less susceptible to  
CC cleavage by proteases and exopeptidases. As these sequences activate p53  
CC DNA binding, they can be used to identify p53 mutants. The peptides can  
CC also be used for treating a patient with a tumour expressing a p53 mutant  
CC whose ability to bind DNA may be activated by one of the peptides. They  
CC can also be used for treating conditions such as exposure to DNA damaging  
CC agents, abnormal cell proliferation characteristic of psoriasis,  
CC atherosclerosis, cancer, arterial restenosis, autoimmune diseases and  
CC undesirable immune responses accompanying rejection of a transplant. The  
CC peptides can also induce apoptosis of specific cells, such as  
CC proliferating lymphocytes.  
SQ Sequence 27 AA:

Query Match 100.0%; Score 58; DB 21; Length 27;  
Best Local Similarity 100.0%; Pred. No. 4,96e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 3 gsrashshl 11  
OY 1 GSRASHSHL 9  
IIIIIIII

RESULT 6  
ID W05358 standard; peptide: 27 AA.  
AC W05358:  
DT 30-APR-1997 (first entry)  
DE Residues 360-386 of p53.  
KW Human; p53: cell proliferation; cell death; regulator; tumour; psoriasis;  
KW negative regulatory region; DNA damaging agent; transplant rejection;  
KW abnormal cell proliferation; atherosclerosis; cancer; autoimmune disease;  
KW arterial restenosis; immune response; apoptosis; inducer; therapy;  
KW proliferating lymphocytes.  
OS Homo sapiens.  
PN W09625434-A1.  
PD 22-AUG-1996.  
PF 16-FEB-1996: U01535.  
PR 16-FEB-1995: US-392542.  
PA (FARB ) BAYER CORP.  
PA (WIST-) WISTAR INST.  
PI Halazonetis T, Hartwig W;  
DR WPI: 96-393345/39.  
PT New human p53-isomorph peptide(s) and peptide:mimetic cpds. - used  
PT for activating p53 function, e.g. for treating tumours, cancers,  
PT psoriasis, etc  
PS Claim 2: Page 37: 55pp: English.  
CC W05350-W05364 represent the p53 (see W05344 for full length wild type  
CC sequence) peptides of the invention. The p53 protein functions to  
CC regulate cell proliferation and cell death, and is mutated in more than  
CC half of all human tumours. These sequences are used to activate the DNA  
CC binding activity of wild type p53, and p53 mutants (see W05345-W05349).  
CC The peptides of the invention consist of at least four sequential amino  
CC acids from a negative regulatory region which maps to residues 361-383 of

CC p53. These sequences preferably contain four amino acids from a non-human  
CC p53 sequence, contain D-form amino acids, and can also be cyclic  
CC peptides. The sequences retain the structural characteristics of the  
CC original peptides, but the modifications render them less susceptible to  
CC cleavage by proteases and exopeptidases. As these sequences activate p53  
CC DNA binding, they can be used to identify p53 mutants. The peptides can  
CC also be used for treating a patient with a tumour expressing a p53 mutant  
CC whose ability to bind DNA may be activated by one of the peptides. They  
CC can also be used for treating conditions such as exposure to DNA damaging  
CC agents, abnormal cell proliferation characteristic of psoriasis,  
CC atherosclerosis, cancer, arterial restenosis, autoimmune diseases and  
CC undesirable immune responses accompanying rejection of a transplant. The  
CC peptides can also induce apoptosis of specific cells, such as  
CC proliferating lymphocytes.  
SQ Sequence 27 AA:

Query Match 100.0%; Score 58; DB 21; Length 27;  
Best Local Similarity 100.0%; Pred. No. 4,96e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 2 gsrashshl 10  
OY 1 GSRASHSHL 9  
IIIIIIII

RESULT 7  
ID W05366 standard; peptide: 27 AA.  
AC W05366:  
DT 30-APR-1997 (first entry)  
DE Peptide p53p360-386DG.  
KW Human; p53: cell proliferation; cell death; regulator; tumour; psoriasis;  
KW negative regulatory region; DNA damaging agent; transplant rejection;  
KW abnormal cell proliferation; atherosclerosis; cancer; autoimmune disease;  
KW arterial restenosis; immune response; apoptosis; inducer; therapy;  
KW proliferating lymphocytes.  
OS Synthetic.  
PN W09625434-A1.  
PD 22-AUG-1996.  
PF 16-FEB-1996: U01535.  
PR 16-FEB-1995: US-392542.  
PA (FARB ) BAYER CORP.  
PA (WIST-) WISTAR INST.  
PI Halazonetis T, Hartwig W;  
DR WPI: 96-393345/39.  
PT New human p53-isomorph peptide(s) and peptide:mimetic cpds. - used  
PT for activating p53 function, e.g. for treating tumours, cancers,  
PT psoriasis, etc  
PS Disclosure: Page 12: 55pp: English.  
CC W05365-W05374 represent examples of the p53 (see W05344 for full length  
CC wild type sequence) peptides of the invention. These sequences all have  
CC additions or deletions of residues from the wild type peptide fragments  
CC of the invention (see W05350-W05364). The p53 protein functions to  
CC regulate cell proliferation and cell death, and is mutated in more than  
CC half of all human tumours. These sequences are used to activate the DNA  
CC binding activity of wild type p53, and p53 mutants (see W05345-W05349).  
CC The peptides of the invention consist of at least four sequential amino  
CC acids from a negative regulatory region which maps to residues 361-383 of  
CC p53. These sequences preferably contain four amino acids from a non-human  
CC p53 sequence, contain D-form amino acids, and can also be cyclic  
CC peptides. The sequences retain the structural characteristics of the  
CC original peptides, but the modifications render them less susceptible to  
CC cleavage by proteases and exopeptidases. As these sequences activate p53  
CC DNA binding, they can be used to identify p53 mutants. The peptides can  
CC also be used for treating a patient with a tumour expressing a p53 mutant  
CC whose ability to bind DNA may be activated by one of the peptides. They  
CC can also be used for treating conditions such as exposure to DNA damaging  
CC agents, abnormal cell proliferation characteristic of psoriasis,  
CC atherosclerosis, cancer, arterial restenosis, autoimmune diseases and  
CC undesirable immune responses accompanying rejection of a transplant. The  
CC peptides can also induce apoptosis of specific cells, such as  
CC proliferating lymphocytes.  
SQ Sequence 27 AA:

Query Match 100.0%; Score 58; DB 21; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 4.96e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 gsrshsahl 10  
 |||||  
 OY 1 GSRHSHSL 9

RESULT 8  
 ID W05365 standard; peptide; 28 AA.

PT 30-APR-1997 (first entry)  
 DE Peptide p53pc360-386.  
 KW Human: p53; cell proliferation; cell death; regulator; tumour; psoriasis;  
 KW negative regulatory region; DNA damaging agent; transplant rejection;  
 KW abnormal cell proliferation; atherosclerosis; cancer; autoimmune disease;  
 KW arterial restenosis; immune response; apoptosis; inducer; therapy;  
 KW proliferating lymphocytes.  
 OS Synthetic.  
 PN W09625434-A1.  
 PD 22-AUG-1996.  
 PE 16-FEB-1996; 001535.  
 PR 16-FEB-1996; US-392542.  
 PA (FARB ) BAYER CORP.  
 PA (WIST-) WISTAR INST.  
 PI Halazoneis T, Hartwig W;  
 DR WPI; 96-39345/39.  
 PT New human p53-isomorphous peptide(s) and peptide-mimetic cpds. - used  
 PT for activating p53 function, e.g. for treating tumours, cancers,  
 PT psoriasis, etc  
 PS Disclosure: Page 12: 55pp; English.  
 CC W05365-W05374 represent examples of the p53 (see W05344 for full length  
 CC wild type sequence) peptides of the invention. These sequences all have  
 CC additions or deletions of residues from the wild type peptide fragments  
 CC A, the invention (see W05350-W05364). The p53 protein functions to  
 CC regulate cell proliferation and cell death, and is mutated in more than  
 CC half of all human tumours. These sequences are used to activate the DNA  
 CC binding activity of wild type p53, and p53 mutants (see W05345-W05349).  
 CC The peptides of the invention consist of at least four sequential amino  
 CC acids from a negative regulatory region which maps to residues 361-383 of  
 CC p53. These sequences preferably contain four amino acids from a non-human  
 CC p53 sequence, contain D-form amino acids, and can also be cyclic  
 CC peptides. The sequences retain the structural characteristics of the  
 CC original peptides, but the modifications render them less susceptible to  
 CC cleavage by proteases and exopeptidases. As these sequences activate p53  
 CC DNA binding, they can be used to identify p53 mutants. The peptides can  
 CC also be used for treating a patient with a tumour expressing a p53 mutant  
 CC whose ability to bind DNA may be activated by one of the peptides. They  
 CC can also be used for treating conditions such as exposure to DNA damaging  
 CC agents, abnormal cell proliferation characteristic of psoriasis,  
 CC atherosclerosis, cancer, arterial restenosis, autoimmune diseases and  
 CC undesirable immune responses accompanying rejection of a transplant. The  
 CC peptides can also induce apoptosis of specific cells, such as  
 CC proliferating lymphocytes.  
 CC Sequence 28 AA:  
 SQ

Query Match 100.0%; Score 58; DB 21; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 4.96e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 gsrshsahl 11  
 |||||  
 OY 1 GSRHSHSL 9

RESULT 9

ID W09322 standard; peptide; 74 AA.  
 AC W09322;  
 DT 10-JUN-1997 (first entry)  
 DE C-terminal domain of p53 protein.  
 KW Chimeric; bispecific; DNA binding domain; trans; activator; repressor;  
 KW diptheria; Pseudomonas; toxin; thymidine kinase; single chain antibody;

KW pathogen; HIV Tat; papilloma virus; E6/E7; Epstein-Barr virus; EBNA;  
 KW hyperproliferation; p53; tumour; oligomerisation.  
 OS Homo sapiens.  
 PN W09630512-A1.  
 PD 03-OCT-1996.  
 PE 29-MAR-1996; F00477.  
 PR 31-MAR-1995; FR-003841.  
 PA (RHON ) RHONE-POULENC RORER SA.  
 PI Biacco L, Schweighoffer F, Toocque B;  
 DR WPI; 96-453539/45.

PT Conditional gene expression system triggered by e.g. infection or  
 PT hyperproliferation - comprises novel bi-specific proteins having  
 PT DNA-binding domain and second domain specific for trans-activator or  
 PT repressor, for gene therapy  
 PS Claim 16; Page 44; 81pp; French.  
 CC The invention relates to novel chimeric, bispecific proteins which  
 CC comprise: (a) a DNA binding domain and (b) a domain which binds a  
 CC trans-activator (TA), trans-repressor (TR) or their complexes, which are  
 CC characteristic of a physiological or pathological state. The novel  
 CC chimeric, bispecific proteins allow expression of a therapeutic protein  
 CC (e.g. diptheria or Pseudomonas toxins, thymidine kinase, single chain  
 CC antibodies) to be regulated in response to particular conditions.  
 CC Examples include making the protein responsive to the presence of  
 CC particular pathogenic TA mols (e.g. HIV Tat, papilloma virus E6/E7  
 CC proteins or Epstein-Barr virus EBNA protein), the therapeutic protein  
 CC will be expressed in those cells infected by that pathogen. Similarly,  
 CC where the chimeric protein responds to a cellular protein typical of a  
 CC hyperproliferative state (esp. wild-type and mutant p53), expression can  
 CC be restricted to tumour cells. The sequence presented here is an example  
 CC of a TA binding domain. It corresponds to the C-terminal domain of the  
 CC p53 protein between residues 320-393 containing the oligomerisation  
 CC domain which binds TA proteins.  
 CC Sequence 74 AA:  
 SQ

Query Match 100.0%; Score 58; DB 21; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 4.96e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 42 gsrshsahl 50  
 |||||  
 OY 1 GSRHSHSL 9

RESULT 10  
 ID R51878 standard; Protein; 157 AA.  
 AC R51878;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 237-393.  
 KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 KW anti-oncogene; cancer; tumour; antibody binding region; epitope.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc.difference 37 /note= "Arg corresponds to a CAT codon"  
 FT PN W09408241-A.  
 PD 14-APR-1994.  
 PE 30-SEP-1993; E02666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schanz P, Tesserer C, Volkman M, Zentgraf H;  
 DR WPI; 94-135732/16.  
 DE N-PSDB; Q62363.  
 PT Non-radioactive detection of p53 specific antibodies - by capture  
 PT on immobilised p53 or its fragments, then reaction with labelled  
 PT second antibody, for diagnosis of tumours and suitable for  
 PT screening  
 PS Claim 10; Page 19: 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 CC Sequence 157 AA:  
 SQ

Query Match 100.0%; Score 58; DB 10; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 4,96e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 125 gsrashshl 133  
 1 GSRASHSHL 9

QY

RESULT 11  
 ID R51876 standard; Protein: 328 AA.  
 AC R51876;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 66-393.  
 KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 OS anti-oncogene; cancer; tumour; antibody binding region; epitope.  
 FH Homo sapiens.  
 Key Location/Qualifiers  
 FT misc\_difference 208  
 FT /note= "Arg corresponds to a CAT codon"  
 PN W09408241-A.  
 PD 14-APR-1994.  
 PE 30-SEP-1993; E02666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schranz P, Tessmer C, Volkmann W, Zentgraf H;  
 DR WPI: 94-135732/16.  
 N-PSDB: Q62361.  
 PT Non-radioactive detection of p53 specific antibodies - by capture  
 on immobilised p53 or its fragments, then reaction with labelled  
 PT second antibody, for diagnosis of tumours and suitable for  
 screening  
 PS Claim 10; Page 18; 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 SQ Sequence 328 AA;

Query Match 100.0%; Score 58; DB 10; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 4,96e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 296 gsrashshl 304  
 1 GSRASHSHL 9

QY

RESULT 12  
 ID W28493 standard; Protein: 353 AA.  
 AC W28493;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 393-325 encoded by p53L177.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 PN W09704092-A1.  
 PD 06-FEB-1997.  
 PE 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON) RHONE-POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 N-PSDB: T66222.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 37; Pages 90-92; 133pp; French.

CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-393 of p53. The present sequence is that of  
 CC a specifically claimed p53 variant designated 393-325 and comprising  
 CC the 325-393 domain, amino acids 75-325 of human wild-type p53 and a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 353 AA;

Query Match 100.0%; Score 58; DB 24; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 4,96e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 38 gsrashshl 46  
 1 GSRASHSHL 9

QY

RESULT 13  
 ID W28494 standard; Protein: 353 AA.  
 AC W28494;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 393-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_difference 179  
 FT /note= "Arg residue at position 182 of wild-type  
 p53 has been mutated to His"  
 PN W09704092-A1.  
 PD 06-FEB-1997.  
 PE 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON) RHONE-POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 37; Page -; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-393 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 393-325H and comprising  
 CC the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 393-325).  
 SQ Sequence 353 AA;

Query Match 100.0%; Score 58; DB 24; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 4,96e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 38 gsrashshl 46

```

OY          1 GSRASHSHL 9
|||||
RESULT 14
ID      R51874 standard; Protein; 354 AA.
AC      R51874;
DE      18-NOV-1994 (first entry)
KW      Human p53 amino acids 40-393.
KW      Human nuclear phosphoprotein p53; tumour suppressor gene product;
KW      anti-oncogene; cancer; tumour; antibody binding region; epitope.
FT      Homo sapiens
FT      Key
FT      misc_difference 234
FT      Location/Qualifiers
FT      Note- "Arg corresponds to a CAT codon"
PN      W09408241 A.
PD      14-APR-1994.
PE      30-SEP-1993; E02666.
PR      30-SEP-1992; DE-232823.
PA      (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.
PI      Klein R, Schranz P, Tessmer C, Volkmann M, Zentgraf H;
DR      WPI: 94-135732/16.
DR      N-PSDB: 062359.
PT      Non-radioactive detection of p53 specific antibodies - by capture
PT      on immobilised p53 or its fragments, then reaction with labelled
PT      second antibody, for diagnosis of tumours and suitable for
PT      screening
PS      Claim 10; Page 18; 35pp; German.
CC      Antibodies specific for p53 are detected by binding to immobilised
CC      fragments of the p53 gene product containing the antibody-binding
CC      region. Preferred fragments contain amino acids 1-241, 40-349,
CC      40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or
CC      368-386. See R51872-R51881 for sequences of these fragments.
SQ      Sequence 354 AA;

Query Match          100.0%; Score 58; DB 10; Length 354;
Best Local Similarity 100.0%; Pred. No. 4,96e-01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      322 gsrashshl 330
OY      1 GSRASHSHL 9
|||||

```

```

RESULT 15
ID      W02617 standard; Protein; 393 AA.
AC      W02617;
DE      06-NOV-1996 (first entry)
DE      Human p53 tumour suppressor protein.
KW      p53 protein; tumour suppressor; tetramerisation domain;
KW      chimaeric protein; gene therapy; vector; cell proliferation; cancer;
KW      apoptosis; autoimmune disease; immune tolerance.
OS      Homo sapiens.
FT      Key
FT      Location/Qualifiers
FT      domain
FT      1..90
FT      /label- Transcription_activation_domain
FT      domain
FT      90..289
FT      /label- DNA-binding_domain
FT      region
FT      316..325
FT      /label- Nuclear_localisation_signal
FT      domain
FT      322..355
FT      /label- Tetramerisation_domain
FT      domain
FT      364..393
FT      /label- Regulation_of_DNA_binding_domain
FT      region
FT      369..375
FT      /label- Nuclear_localisation_signal
FT      region
FT      379..384
FT      /label- Nuclear_localisation_signal
PN      W09616989-A1.
PD      06-JUN-1996.
PE      27-NOV-1995; U15353.
PR      28-NOV-1994; US-347792.
PR      28-APR-1995; US-431357.

```

```

PR      01-JUN-1995; US-456623.
PA      (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI      Halazoneis TD.
DR      WPI: 96-286828/29.
DR      N-PSDB: T32831.
PT      New chimaeric p53 protein with heterologous tetramerisation domain
PT      - and related DNA and vectors, useful for treating abnormal cell
PT      proliferation, esp. cancer, auto-immune disease, etc.
PS      Disclosure, Page 68-70; 123pp; English.
CC      Human wild-type p53 (W02617) is a sequence-specific DNA binding
CC      protein with tumour suppressor function. It regulates cell
CC      proliferation and apoptosis and participates in cellular response
CC      to DNA damaging agents. It is inactivated in more than half of
CC      all human tumours. Novel chimaeric p53 proteins have altered
CC      tetramerization domains (see also W02622), retain wild-type p53
CC      function and form tetramers, but do not hetero-oligomerise with
CC      wild-type p53 or tumour-derived p53 mutants, and/or have restricted
CC      DNA binding specificity. These proteins, and nucleic acids encoding
CC      them, can be used to treat abnormal cell proliferation, esp. cancer,
CC      or to induce immune tolerance to facilitate transplants and treat
CC      autoimmune diseases.
SQ      Sequence 393 AA;

Query Match          100.0%; Score 58; DB 19; Length 393;
Best Local Similarity 100.0%; Pred. No. 4,96e-01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      361 gsrashshl 369
OY      1 GSRASHSHL 9
|||||

```

```

Search completed: Fri Sep 11 13:56:06 1998
Job time : 13 secs.

```

\*\*\*\*\*  
 WISE (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Sep 11 13:56:24 1998; MasPar time 3.18 Seconds  
 Tabular output not generated. 103.412 Million cell updates/sec

Title: >US-08-452-843-21  
 Description: (1-9) from US08452843.pep  
 Perfect Score: 58  
 Sequence: 1 GSRHSHSL 9

Scoring table: PAM 150  
 Gap 15

Searched: 120441 seqs, 3653193 residues  
 Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: p1r56  
 1:p1r1 2:p1r2 3:p1r3 4:p1r4 5:nr13d

Statistics: Mean 19.650; Variance 22.450; scale 0.875

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	58	100.0	393	1	DNHU53	1.82e-03
2	58	100.0	393	2	S05594	1.82e-03
3	51	87.9	386	2	S51648	1.10e-01
4	49	84.5	391	2	JC6193	3.39e-01
5	46	79.3	143	2	S52595	1.75e+00
6	43	74.1	201	1	OKHU2	8.44e+00
7	42	72.4	97	2	S01502	1.41e+01
8	42	72.4	167	2	S25360	1.41e+01
9	42	72.4	259	2	B69933	1.41e+01
10	42	72.4	384	2	JQ2351	1.41e+01
11	42	72.4	390	1	DNMS53	1.41e+01
12	42	72.4	998	2	S37627	1.41e+01
13	42	72.4	3268	2	S69625	1.41e+01
14	41	70.7	49	5	IFSU1	1.41e+01
15	41	70.7	371	1	TVXLT1	2.32e+01
16	41	70.7	533	1	KJHUB	2.32e+01
17	41	70.7	688	2	A44306	2.32e+01
18	41	70.7	1018	2	JC5799	2.32e+01
19	41	70.7	1534	2	A56734	2.32e+01
20	40	68.0	77	2	S29563	2.32e+01
21	40	68.0	97	2	D34284	3.80e+01
22	40	69.0	151	2	S48796	3.80e+01
23	40	69.0	287	2	S71548	3.80e+01

24	40	69.0	414	2	B64033	3.80e+01
25	40	69.0	446	2	S59646 <td>3.80e+01</td>	3.80e+01
26	40	69.0	456	2	S66080 <td>3.80e+01</td>	3.80e+01
27	40	69.0	456	2	S05371 <td>3.80e+01</td>	3.80e+01
28	40	69.0	533	2	S18539 <td>3.80e+01</td>	3.80e+01
29	40	69.0	598	2	S39621 <td>3.80e+01</td>	3.80e+01
30	40	69.0	600	2	B46642 <td>3.80e+01</td>	3.80e+01
31	40	69.0	621	2	S73155 <td>3.80e+01</td>	3.80e+01
32	40	69.0	959	2	S61155 <td>3.80e+01</td>	3.80e+01
33	40	69.0	1379	2	S01254 <td>3.80e+01</td>	3.80e+01
34	40	69.0	1390	1	TYHME <td>3.80e+01</td>	3.80e+01
35	40	69.0	1748	2	JW0786 <td>3.80e+01</td>	3.80e+01
36	39	67.2	66	2	S01285 <td>6.17e+01</td>	6.17e+01
37	39	67.2	234	2	A55367 <td>6.17e+01</td>	6.17e+01
38	39	67.2	414	2	A56419 <td>6.17e+01</td>	6.17e+01
39	39	67.2	419	2	B49418 <td>6.17e+01</td>	6.17e+01
40	39	67.2	427	2	S54574 <td>6.17e+01</td>	6.17e+01
41	39	67.2	457	2	JC6026 <td>6.17e+01</td>	6.17e+01
42	39	67.2	496	2	A49418 <td>6.17e+01</td>	6.17e+01
43	39	67.2	1375	2	JC5148 <td>6.17e+01</td>	6.17e+01
44	39	67.2	1706	2	S51499 <td>6.17e+01</td>	6.17e+01
45	39	67.2	3461	2	S58870 <td>6.17e+01</td>	6.17e+01

## ALIGNMENTS

RESULT 1

ENTRY DNHU53 #type complete  
 TITLE cellular tumor antigen p53 - human  
 ALTERNATE\_NAMES cellular phosphoprotein p53; oncoprotein p53; transformation suppressor p53; tumor suppressor p53  
 ORGANISM Homo sapiens #common\_name man  
 DATE 05-Oct-1988 #sequence\_revision 18-Nov-1994 #text\_change 18-Sep-1997

ACCESSIONS A25224; A43073; J70436; S40773; S42669; A22837; A55060; A25397; B25397; S42453; S42453; S38082; S38083; S38084; S38085; S38086; S38087; S38088; S38089; S38090; S38091; S38092; S38093; A44905; I58354; I78850; S60153

REFERENCE A25224  
 #authors Lamb, P.; Crawford, L.  
 #journal Mol. Cell. Biol. (1986) 6:1379-1385  
 #title Characterization of the human p53 gene.  
 #cross-references MIMD:87064416  
 #accession A25224  
 #molecule\_type DNA  
 #residues 1-393 #label IAM  
 #cross-references EMBL:X01405; GB:M13121; GB:N00092; NID:g189460; PDB:g386994

REFERENCE J70436  
 #authors Buchman, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.; Georgiev, G.P.  
 #journal Gene (1988) 70:245-252  
 #title A variation in the structure of the protein-coding region of the human p53 gene.  
 #cross-references MIMD:89108008  
 #accession A43073  
 #molecule\_type DNA  
 #residues 1-393 #label BUC  
 #residues this 72-Arg allele appears to be about 5 times more frequent than the 72-Pro allele

#accession J70436  
 #molecule\_type DNA  
 #residues 1-71, 'P', 73-393 #label BUC  
 #cross-references EMBL:M22898; NID:g189474; PDB:g189476  
 #note this 72-Pro allele was found in both normal and malignant cell lines

REFERENCE S40773  
 #authors Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.  
 #submission submitted to the EMBL Data Library, August 1990  
 #accession S40773  
 #molecule\_type DNA  
 #residues 1-393 #label CHU  
 #cross-references EMBL:X54156; NID:g35213; PDB:g35214



REFERENCE  
#authors S42669  
#journal Matlashewski, G.; Lamb, P.; Pim, D.; Peacock, J.; Crawford,  
L.; Benchimol, S.  
#title EMBO J. (1984) 3:3257-3262  
#note Isolation and characterization of a human p53 cDNA clone:  
expression of the human p53 gene.  
#accession S42669  
##molecule-type mRNA  
##residues 101-393 #label MK1  
#cross-references EMBL:X01405; NID:g35215; PID:g642241  
REFERENCE  
#authors A22837  
#journal Zakut-Houri, R.; Biernz-Tadnor, B.; Givol, D.; Oren, M.  
#title EMBO J. (1985) 4:1251-1255  
#note Human p53 cellular tumor antigen: cDNA sequence and  
expression in COS cells.  
#cross-references MIMD:85230577  
#accession A22837  
##molecule-type mRNA  
##residues 1-71, 'P', '73-393 #label ZAK  
#cross-references EMBL:X02469; EMBL:M60950; NID:g35209; PID:g35210  
REFERENCE  
#authors A55060  
#journal Harlow, E.; Williamson, N.M.; Raiston, R.; Helfman, D.M.;  
Adams, T.E.  
#title Mol. Cell. Biol. (1985) 5:1601-1610  
#note Molecular cloning and in vitro expression of a cDNA clone for  
human cellular tumor antigen p53.  
#accession A55060  
##molecule-type mRNA  
##residues 1-71, 'P', '73-272, 'H', '274-393 #label HA3  
#cross-references GB:K03199; NID:g189478; PID:g189479  
#experimental-source clone pr4-2, cell line A431  
REFERENCE  
#authors A93086  
#journal Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.;  
Arai, N.; Rotter, V.  
#title Mol. Cell. Biol. (1986) 6:4650-4656  
#note Molecular basis for heterogeneity of the human p53 protein.  
#cross-references MIMD:87089826  
#accession A25397  
##molecule-type mRNA  
##residues 1-78, 'T', '80-393 #label HAR  
#cross-references EMBL:M14694; NID:g339613; PID:g339614  
#experimental-source clone p53-H-1, transformed hybridoma SV-80 cell  
line  
#accession B25397  
##molecule-type mRNA  
##residues 1-71, 'P', '73-78, 'T', '80-393 #label HA2  
#cross-references EMBL:M14695; NID:g339615; PID:g339616  
#experimental-source clone p53-H-19, transformed hybridoma SV-80 cell  
line  
REFERENCE  
#authors S42452  
#journal Matlashewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider,  
J.; Crawford, L.V.  
#title Mol. Cell. Biol. (1987) 7:961-963  
#note Primary structure polymorphism at amino acid residue 72 of  
human p53.  
#accession S42452  
##molecule-type mRNA  
##residues 66-71, 'P', '73-79 #label MK2  
#experimental-source clone lambda C113  
#note 72-Cys was also found, and appears to represent a  
polymorphism  
#accession S42453  
##molecule-type mRNA  
##residues 66-79 #label MAT  
#experimental-source clone J6K  
REFERENCE  
#authors I38082  
#journal Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.;  
Crook, T.  
#title EMBO J. (1991) 10:2879-2887  
#note p53 is frequently mutated in Burkitt's lymphoma cell lines.  
#cross-references MIMD:92007731  
#accession I38082  
#status translated from GB/EMBL/DBJ

##molecule-type mRNA  
##residues 1-189, 'L', '189-393 #label F01  
#cross-references EMBL:X60010; NID:g506433; PID:g506433  
#note deletion of a C nucleotide causes a frameshift at  
position 566  
#accession I38083  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-192, 'R', '194-393 #label F02  
#cross-references EMBL:X60011; NID:g506434; PID:g506435  
#accession I38084  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-393 #label F03  
#cross-references EMBL:X60012; NID:g506436; PID:g506437  
#accession I38085  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-245, 'T', '247-393 #label F04  
#cross-references EMBL:X60013; NID:g506438; PID:g506439  
#accession I38086  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-236, 'T', '238-393 #label F05  
#cross-references EMBL:X60014; NID:g506440; PID:g506441  
#accession I38087  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-247, 'Q', '249-393 #label F06  
#cross-references EMBL:X60015; NID:g506442; PID:g506443  
#accession I38088  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-71, 'P', '73-237, 'Y', '239-393 #label F07  
#cross-references EMBL:X60016; NID:g506444; PID:g506445  
#accession I38089  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-247, 'Q', '249-393 #label F08  
#cross-references EMBL:X60017; NID:g506446; PID:g506447  
#accession I38090  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-71, 'P', '73-162, 'H', '164-393 #label F09  
#cross-references EMBL:X60018; NID:g506448; PID:g506449  
#accession I38091  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-212, 'Q', '214-393 #label F10  
#cross-references EMBL:X60019; NID:g506450; PID:g506451  
#accession I38092  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-253, 'D', '255-393 #label F11  
#cross-references EMBL:X60020; NID:g506452; PID:g506453  
#note all sequences submitted to the EMBL/Genbank/DBJ  
databases June 1991  
REFERENCE  
#authors I38093  
#journal Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.  
#title Nucleic Acids Res. (1991) 19:6977  
#note An Alu polymorphism intragenic to the TP53 gene.  
#cross-references MIMD:92107726  
#accession I38093  
##status translated from GB/EMBL/DBJ  
##molecule-type DNA  
##residues 1-393 #label RE2  
#cross-references EMBL:X54156; NID:g35213; PID:g35214  
REFERENCE  
#authors A44905  
#journal Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;  
Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.;  
Teraoka, M.  
#title Cancer Res. (1991) 51:5800-5805



#title p53 gene mutations in gastric cancer metastases and in  
gastric cancer cell lines derived from metastases.  
#cross-references MIMD:92034678  
#accession A44905  
##molecule-type DNA  
##residues 246-247, 'W', 249-250 ##label YAM  
##cross-references GB:S63157; NID:9237829; PID:9237830  
#note sequence extracted from NCBI backbone (NCBIN:63157,  
Note: remainder of annotations omitted.

Query Match 1 100.0%; Score 58; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.82e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 361 GSRASHSHL 369  
1 GSRASHSHL 9

RESULT 2  
ENTRY S06594 #type complete  
TITLE cellular tumor antigen p53 - green monkey  
ORGANISM #formal\_name Cercopithecus aethiops #common\_name green  
monkey, grivet  
DATE 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change  
08-Sep-1997

ACCESSIONS S06594  
REFERENCE S06594  
#authors Rigaudy, P.; Eckhart, W.  
#journal Nucleic Acids Res. (1989) 17:8375  
#title Nucleotide sequence of a cDNA encoding the monkey cellular  
phosphoprotein p53.  
#cross-references MIMD:90045967  
#accession S06594  
##molecule-type mRNA  
##residues 1-393 ##label RIG  
#cross-references EMBL:X16384; NID:922795; PID:922796

CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;  
nucleus; phosphoprotein; transcription regulation; tumor  
suppressor; zinc

FEATURE  
176,179,238,242 #binding\_site zinc (Cys, His, Cys, Cys) #status  
predicted  
392 #binding\_site phosphoryl-RNA (Ser) (covalent) #status  
predicted

SUMMARY #length 393 #molecular-weight 43696 #checksum 4263  
Query Match 100.0%; Score 58; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.82e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 361 GSRASHSHL 369  
1 GSRASHSHL 9

RESULT 3  
ENTRY S51648 #type complete  
TITLE cellular tumor antigen p53 - bovine  
ALTERNATE\_NAMES tumor-suppressor protein p53  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change  
08-Sep-1997  
ACCESSIONS S51648  
REFERENCE S51648  
#authors Deguict, F.; Williams, L.; Burny, A.; Kettmann, R.  
#submission submitted to the EMBL Data Library, September 1994  
#description Nucleotide sequence of the ovine p53 tumor-suppressor gene  
cDNA and its genomic organisation.  
#accession S51648  
#status preliminary

##molecule-type mRNA  
##residues 1-386 ##label DEO  
##cross-references EMBL:X81704; NID:9602332; PID:9602333  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;  
phosphoprotein; transcription regulation; tumor suppressor;  
zinc

FEATURE  
166,171,231,235 #binding\_site zinc (Cys, His, Cys, Cys) #status  
predicted  
385 #binding\_site phosphoryl-RNA (Ser) (covalent) #status  
predicted

SUMMARY #length 386 #molecular-weight 43255 #checksum 7025  
Query Match 87.9%; Score 51; DB 2; Length 386;  
Best Local Similarity 100.0%; Pred. No. 1.10e-01;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 355 SRAHSHL 362  
2 SRAHSHL 9

RESULT 4  
ENTRY JC6193 #type complete  
TITLE tumor suppressor p53 - rabbit  
ORGANISM #formal\_name Oryctolagus cuniculus #common\_name domestic  
rabbit  
DATE 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change  
08-Sep-1997

ACCESSIONS JC6193  
REFERENCE JC6193  
#authors Le Goas, F.; May, P.; Ronco, P.; de Fromental, C.C.  
#journal Gene (1997) 185:169-173  
#title cDNA cloning and immunological characterization of rabbit  
p53.  
#accession JC6193  
##molecule-type mRNA  
##residues 1-391 ##label LEA  
#cross-references EMBL:X90592; NID:91532043; PID:9194962; PID:91532044

GENETICS  
#gene p53  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS tumor  
SUMMARY #length 391 #molecular-weight 43435 #checksum 4367

Query Match 84.5%; Score 49; DB 2; Length 391;  
Best Local Similarity 88.9%; Pred. No. 3.39e-01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 359 GSRASHSHL 367  
1 GSRASHSHL 9

RESULT 5  
ENTRY S52595 #type complete  
TITLE probable membrane protein YHR056w-a - yeast (Saccharomyces  
cerevisiae)  
ORGANISM #formal\_name Saccharomyces cerevisiae  
DATE 05-May-1995 #sequence\_revision 19-Oct-1995 #text\_change  
21-Nov-1997  
ACCESSIONS S52595  
REFERENCE S46729  
#authors Du, Z.  
#submission submitted to the EMBL Data Library, May 1994  
#description The sequence of S. cerevisiae cosmid 8025.  
#accession S52595  
##molecule-type DNA  
##residues 1-143 ##label DUZ  
#cross-references EMBL:U00061; MIPS:YHR056w-a  
GENETICS  
#map\_position 8R

## KEYWORDS transmembrane protein

FEATURE 15-31 #domain transmembrane #status predicted #label TM  
SUMMARY #length 143 #molecular-weight 16054 #checksum 8917

Query Match 79.3%; Score 46; DB 2; Length 143;  
Best Local Similarity 66.7%; Pred. No. 1.75e+00;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 85 GRRGSHDL 93

QY 1 GSRAHSHL 9

RESULT 6  
ENTRY OMH2 #type complete  
TITLE alpha-1-acid glycoprotein 2 precursor - human  
ALTERNATE\_NAMES alpha-1-acid glycoprotein B; orosomucoid 2  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 31-Mar-1992 #sequence\_revision 07-Jun-1996 #text\_change 05-Sep-1997

ACCESSIONS JT0326; B28346  
REFERENCE JT0326

#authors Merritt, C.M.; Board, P.G.  
#journal Gene (1988) 66:97-106  
#title Structure and characterisation of a duplicated human alpha 1 acid-glycoprotein gene.

#cross-references MUID:86329732  
#accession JT0326

#molecule\_type DNA  
#residues 1-201 #label MER

REFERENCE #cross-references GB:M21540; NID:9177839; PID:9177840  
#authors Dente, L.; Pizsa, M.G.; Metspalu, A.; Cortese, R.  
#journal EMBO J. (1987) 6:2289-2296

#title Structure and expression of the genes coding for human alpha-1-acid glycoprotein.  
#cross-references MUID:88029318

#accession B28346

#molecule\_type DNA

#residues 1-118, 'N', 120-201 #label DEN

COMMENT #cross-references GB:X06674

COMMENT Alpha-1-AGP, synthesized in the liver and leucocytes, appears to function in modulating the activity of the immune system during the acute-phase reaction.

COMMENT See also OMH1.

GENETICS

#gene GDB:ORF2  
#cross-references GDB:120251; OMIM:138610

#map\_position 9q32-9q32

#introns 38/3; 86/2; 110/1; 146/1; 180/3

CLASSIFICATION #superfamily lipocalin; lipocalin homology

KEYWORDS acute phase; glycoprotein; leukocyte; liver; plasma; pyroglyutamic acid

FEATURE

1-18 #domain signal sequence #status predicted #label SIG

19-201 #product alpha-1-acid glycoprotein 2 #status predicted

34-183 #domain lipocalin homology #label LIP

19 #modified\_site pyrrolidone carboxylic acid (Gln) (1n

23-165, 90-183 #disulfide\_bonds #status predicted

33,56,72,93,103 #binding\_site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 201 #molecular-weight 23602 #checksum 4589

Query Match 74.1%; Score 43; DB 1; Length 201;

Best Local Similarity 55.6%; Pred. No. 8.44e+00;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 111 GGRHVAHL 119

QY 1 GSRAHSHL 9

RESULT 7

ENTRY S01502 #type complete

TITLE NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4L - sea urchin (Strongylocentrotus purpuratus) mitochondrion (SGC8)

ORGANISM #formal\_name Strongylocentrotus purpuratus

DATE 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 09-Sep-1994

ACCESSIONS S01502

REFERENCE S01499

#authors Jacobs, H.T.; Elliott, D.J.; Math, V.B.; Farquharson, A.

#journal J. Mol. Biol. (1988) 202:185-217

#title Nucleotide sequence and gene organization of sea urchin mitochondrion DNA.

#cross-references MUID:89011951

#accession S01502

#residues 1-97 #label JAC

GENETICS #cross-references EMBL:X12631

#gene nd4L

#genome mitochondrion

#genetic\_code SGC8

#start\_codon ATC

CLASSIFICATION

KEYWORDS

SUMMARY

Query Match 72.4%; Score 42; DB 2; Length 97;  
Best Local Similarity 75.0%; Pred. No. 1.41e+01;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 81 SRRHSHL 88

QY 2 SRAHSHL 9

RESULT 8

ENTRY S25360 #type complete

TITLE signal recognition particle protein SRP21 - yeast (Saccharomyces cerevisiae)

ALTERNATE\_NAMES protein YKL122c; protein YKL527

ORGANISM #formal\_name Saccharomyces cerevisiae

DATE 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 06-Feb-1998

ACCESSIONS S25360; S37950; S51951

REFERENCE S25357

#authors Colledge, L.; Richard, G.F.; Thierry, A.; Dujon, B.

#journal Yeast (1992) 8:325-336

#title Sequence of a segment of yeast chromosome XI identifies a new mitochondrial carrier, a new member of the G protein family, and a protein with the PAKK motif of the H1 histones.

#accession S25360

#molecule\_type DNA

#residues 1-167 #label COL

#cross-references EMBL:S44213; NID:9254447; PID:9254451

REFERENCE S37938

#authors Jacquier, A.; Legrain, P.; Colledge, L.; Richard, G.F.; Thierry, A.; Dujon, B.

#journal submitted to the Protein Sequence Database, March 1994

#accession S37950

#molecule\_type DNA

REFERENCE #residues 1-167 #label JAC

#cross-references EMBL:Z28122; NID:9486205; PID:9486206; MIPS:YKL122c

#authors

Brown, J.D.; Hann, B.C.; Medzhradszky, K.F.; Niva, M.; Burlingame, A.L.; Walter, P.

EMBO J. (1994) 13:4390-4400

```

#title      Subunits of the Saccharomyces cerevisiae signal recognition
#accession  S51951
#molecule_type protein
#residues   2-15;45-55;75-85;107-120 ##label BRO
GENETICS
#gene       SGD:SRP21
#cross-references SGD:S0001605; MIPS:YKL122c
#map_position 11L
SUMMARY     #length 167 #molecular-weight 18425 #checksum 4127

Query Match      72.4%; Score 42; DB 2; Length 167;
Best Local Similarity 71.4%; Pred. No. 1.41e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DO 44 RTHNSHL 50
1-1-1111
QY 3 RAHSHSL 9

RESULT 9
ENTRY B69933 #type complete
TITLE  conserved hypothetical protein ypbG - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSIONS B69933
REFERENCE A69380
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Allioni, G.; Azevedo, V.; Bertorello, M.G.; Bessieres, P.;
Bolotin, A.; Borcherdt, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Conterton, J.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, R.M.; Diesterhoef, A.;
Enright, S.D.; Emerson, P.T.; Ertlan, K.D.; Ertling, J.;
Fabre, C.; Ferrari, E.; Foulger, D.; Frit, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen, N.; Ghim,
S.I.; Glaeser, P.; Goffeau, A.; Goldthly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klier-Bianchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Manuel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, T.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Portolik, S.; Prescott,
A.M.; Prescan, E.; Puig, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serron, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vasseroth, A.; Viari, A.;
Wambert, R.; Wedler, E.; Wedler, H.; Weitenegger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#accession B69933
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-259 #label KUN
#experimental_source strain 168
GENETICS

```

```

#gene ypbG
SUMMARY #length 259 #molecular-weight 28560 #checksum 3614

Query Match      72.4%; Score 42; DB 2; Length 259;
Best Local Similarity 50.0%; Pred. No. 1.41e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 68 ARSHAPHL 75
1-1-1-111
QY 2 SRAHSHSL 9

RESULT 10
ENTRY JQ2351 #type complete
TITLE  glycoprotein D precursor - turkey herpesvirus
ALTERNATE_NAMES ORF 6 protein
ORGANISM #formal_name turkey herpesvirus
DATE 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change
ACCESSIONS JQ2351
REFERENCE JQ2346
#authors Zelnik, V.; Dartell, R.; Audonnet, J.C.; Smith, G.D.;
Riviere, M.; Pastorek, J.; Ross, L.J.N.
#journal J. Gen. Virol. (1993) 74:2151-2162
#title The complete sequence and gene organization of the short
unique region of herpesvirus of turkeys.
#accession JQ2351
#molecule_type DNA
#residues 1-384 ##label 2EL
KEYWORDS glycoprotein; transmembrane protein
FEATURE 1-25
26-364
345-360
123,215,220
#domain signal sequence #status predicted #label SIG
#product glycoprotein D #status predicted #label MARV
#domain transmembrane #status predicted #label TMV
#binding site carbohydrate (Asn) (covalent) #status
predicted

SUMMARY #length 384 #molecular-weight 43857 #checksum 4684

Query Match      72.4%; Score 42; DB 2; Length 384;
Best Local Similarity 85.7%; Pred. No. 1.41e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 230 RAHSHSL 236
11111111
QY 3 RAHSHSL 9

RESULT 11
ENTRY DNM553 #type complete
TITLE  cellular tumor antigen p53 - mouse
ALTERNATE_NAMES oncoprotein p53
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Aug-1985 #sequence_revision 04-Oct-1996 #text_change
ACCESSIONS A22739; S06336; A02684; S38822; S38823; I48703
REFERENCE A22739
#authors Blenz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.
#journal EMBO J. (1984) 3:2179-2183
#cross-references MUID:85027173
#accession A22739
#molecule_type DNA
#residues 1-134, 'V', 136-390 #label BIE
REFERENCE S06336
#authors Chumakov, P.M.
#journal Bioorg. Khim. (1987) 13:1691-1694
#title Primary structure of DNA complementary to murine oncoprotein
p53 mRNA.
#cross-references MUID:88221682
#accession S06336
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-134, 'V', 136-390 #label CHU
REFERENCE A02684

```

```

#authors      Zakut-Houri, R.; Oren, M.; Blenz, B.; Lavie, V.; Hazum, S.;
#journal      Nature (1983) 306:594-597
#title        A single gene and a pseudogene for the cellular tumour
#             antigen p53.
#cross-references EMBL:84068204
#accession    A02684
#             ##molecule-type mRNA
#             1-159,'H',161-167,'G',169-233,'T',235-390 ##label ZAK
REFERENCE     S38822
#authors      Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
#             Shohat, O.; Rotter, V.
#journal      Mol. Cell. Biol. (1986) 6:3232-3239
#title        Immunologically distinct p53 molecules generated by
#             alternative splicing.
#accession    S38822
#             ##status      preliminary
#             ##molecule-type mRNA
#             ##residues    1-390 ##label ARA
#cross-references EMBL:M13872; NID:g200198; PID:g200199
#accession    S38823
#             ##status      preliminary
#             ##molecule-type mRNA
#             ##residues    1-167,'G',169-233,'T',235-390 ##label AR2
#cross-references EMBL:M13873
REFERENCE     I48703
#authors      Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.
#journal      Nucleic Acids Res. (1984) 12:5609-5626
#title        Cloning and expression analysis of full length mouse cDNA
#             sequences encoding the transformation associated protein
#             p53.
#cross-references MUID:84272240
#accession    I48703
#             ##status      preliminary; translated from GB/EMBL/DBJ
#             ##molecule-type mRNA
#             ##residues    1-47,'R',49-78,'QW',82-390 ##label RES
#cross-references EMBL:X00741; NID:953570; PID:953571
COMMENT       This DNA-binding protein plays an essential role in the regulation
#             of cell division, as it is required for the transition from phase
#             G0 to G1 of the cell cycle.
COMMENT       The tetramer association region may exhibit a beta-turn,
#             beta-sheet, beta-turn, alpha-helix motif.
#             #superfamily cellular tumor antigen p53
#             #apoptosis; cell division control; DNA binding; homotetramer;
#             phosphoprotein; transcription regulation; tumor suppressor;
#             zinc
CLASSIFICATION
KEYWORDS
FEATURE
1-44          #domain transcription activation #status predicted
              #label TRA\
16-26         #region conserved region I\
99-289        #domain DNA-binding core #status predicted #label DBC\
108-121       #region L1 loop\
114-139       #region conserved region II\
160-192       #region L2 loop\
168-178       #region conserved region III\
231-252       #region conserved region IV\
233-248       #region L3 loop\
267-283       #region conserved region V\
312-319       #region conserved region V\
313-319       #region nuclear location signal\
313-337       #region tetramer association\
7,9,12,18,23,37 #binding-site phosphate (Ser) (covalent) #status
              predicted\
173,176,235,239 #binding-site zinc (Cys, His, Cys, Cys) #status
              predicted\
312           #binding-site phosphate (Ser) (covalent) (by cdc2
              kinase) #status predicted\
389           #binding-site phosphoryl-RNA (Ser) (covalent) #status
              predicted
SUMMARY       #length 390 #molecular-weight 43458 #checksum 1260
Query Match 72.4%; Score 42; DB 1; Length 390;
Best Local Similarity 87.5%; Pred. No. 1,41e+01;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

DB 359 SRAHSSYL 366
|||||
QY 2 SRAHSSHL 9

RESULT 12
ENTRY   S37627 #type complete
TITLE   protein-tyrosine kinase (EC 2.7.1.112), receptor-type - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE    19-May-1994 #sequence_revision 03-Aug-1995 #text_change
19-Dec-1997
ACCESSIONS
REFERENCE
#authors      Boehme, B.; Holtlich, U.; Wolf, G.; Luzius, H.; Grzeschik,
#             K.H.; Strebhardt, K.; Ruedsamen-Waigmann, H.
#journal      Oncogene (1993) 8:2857-2862
#title        PCR mediated detection of a new human
#             receptor-tyrosine-kinase, HEK 2.
#accession    S37627
#             ##status      preliminary
#             ##molecule-type mRNA
#             ##residues    1-998 ##label BOE
#cross-references EMBL:X75208; NID:g406867; PID:g406868
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein
KEYWORDS      kinases; protein kinase homology
FEATURE
631-899      #domain protein kinase homology #label KIM\
633-647      #region protein kinase ATP-binding motif
SUMMARY       #length 998 #molecular-weight 110286 #checksum 4450
Query Match 72.4%; Score 42; DB 2; Length 998;
Best Local Similarity 62.5%; Pred. No. 1,41e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
DB 408 PRVYTRSHL 415
|||||
QY 2 SRAHSSHL 9

RESULT 13
ENTRY   S69625 #type complete
TITLE   hypothetical protein YDR457w - yeast (Saccharomyces
ORGANISM #formal_name Saccharomyces cerevisiae
DATE    22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change
06-Feb-1998
ACCESSIONS
REFERENCE
#authors      Dietrich, F.S.
#submission   submitted to the EMBL Data Library, August 1995
#description   The sequence of S. cerevisiae cosmid 9410, 8035, 8166, and
              9787.
#accession    S69625
#             ##molecule-type DNA
#             ##residues    1-3268 ##label DIE
#cross-references EMBL:U03050; NID:g927726; PID:g927738; MIPS:YDR457w
GENETICS
#gene         SGD:TOM1
#map_position 4R
SUMMARY       #length 3268 #molecular-weight 374181 #checksum 6577
Query Match 72.4%; Score 42; DB 2; Length 3268;
Best Local Similarity 66.7%; Pred. No. 1,41e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
DB 2194 GSRPRSHL 2202
|||||
QY 1 GSRHSSHL 9

```

```

RESULT 14
ENTRY 1FSU1 #type fragment
TITLE n-acetylglactosamine-4-sulfatase (EC 3.1.6.12), fragment 1 - human
ALTERNATE_NAMES 4-sulfatase; arylsulfatase b; asb
PDB_TITLE 4-sulfatase (human)
ORGANISM #formal_name Homo sapiens #common_name man
#note liver, expressed in chinese hamster ovary cells (cho)
#cricketulys griseus, var. kl, ATCC: CCL 61

REFERENCE A68458
#authors Bond, C.; Guss, M.
#submission #submitted to the Brookhaven Protein Data Bank, July 1996
#cross-references PDB:1FSU
TNO39327

#authors Bond, C.S.; Clements, P.R.; Ashby, S.J.; Collyer, C.A.;
#journal Harrop, S.J.; Hopwood, J.J.; Guss, J.M.
#title Structure (London) (1997) 5:277
#comment Resolution: 2.5 angstroms
#comment Determination: X-ray diffraction
#comment R-value: 0.188
#keywords glycosaminoglycan degradation; hydrolase; lysosome; signal
#glycoprotein; sulfatase

FEATURE
20-22 #region helix (right hand 3-10) \
29-36 #region helix (right hand alpha) \
5-11 #region beta sheet \
39-41 #region beta sheet \
12,13 #site Asp, Asp #label MEB
SUMMARY #length 49 #checksum 3519

Query Match 70.7%; Score 41; DB 5; Length 49;
Best Local Similarity 55.6%; Pred. No. 2.32e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 23 GSRRTPHL 31
QY 1 GSRAHSHL 9

RESULT 15
ENTRY TVXLT1 #type complete
TITLE transforming protein int-1 precursor - African clawed frog
ALTERNATE_NAMES pxynt-1 protein
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change
05-Sep-1997
ACCESSIONS S02113; S41630
REFERENCE S02113
#authors Noordermeer, J.; Weijlink, F.; Verrijzer, P.; Rijsewijk, F.;
#journal Destree, O.
#title Nucleic Acids Res. (1989) 17:11-18
#comment Isolation of the Xenopus homolog of int-1/Wingless and
#expression during neurula stages of early development.
#cross-references WUID:89098373
#accession S02113
#molecule_type mRNA
#residues 1-371 #label NOO
REFERENCE S41630
#cross-references EMBL:X13138; NID:965235; PID:965236
#authors Gao, X.; Kuiken, G.A.; Baarends, W.M.; Koster, J.G.; Destree,
#journal O.H.J.
#title Oncogene (1994) 9:573-581
#comment Characterization of a functional promoter for the Xenopus
#int-1 gene in vivo.
#accession S41630
#molecule_type DNA
#residues 1-37 #label GAO
#cross-references EMBL:X56845
GENETICS
#gene int-1
#classification #superfamily int-1 transforming protein
#keywords glycoprotein; oncogene; transforming protein

```

```

FEATURE
1-19 #domain signal sequence #status predicted #label SIG\
20-371 #product transforming protein int-1 #status predicted
#label MAT\
28,261,279,306,317, #binding site carbohydrate (Asn) (covalent) #status
360 predicted
SUMMARY #length 371 #molecular-weight 41125 #checksum 3277

Query Match 70.7%; Score 41; DB 1; Length 371;
Best Local Similarity 55.6%; Pred. No. 2.32e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 267 GSRRDPHL 275
QY 1 GSRAHSHL 9

Search completed: Fri Sep 11 13:56:42 1998
Job time : 18 secs.

```



W.T.

+01

	RESULT	1	ALIGNMENTS
ID	P53-PELCA	STANDARD:	PRT: 366 AA.
AC	P416B5;		
DT	01-NOV-1995 (REL. 32, CREATED)		
DT	01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)		
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)		
DE	CELLULAR TUMOR ANTIGEN P53.		
GN	TP53.		
OS	FELIS SILVESTRIS CATUS (CAT).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;		
CC	ETHERIUM; CARNIVORA.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-LYMPH NODE;		
RX	MEDLINE: 94333960.		
RA	OTUDA M., UMEIDA A., SAKAI T., OHASHI T., MOMOI Y., YOUN H.Y.,		
RA	WATARI T., GOITSUKA R., TSUJIMOTO H., HASEGAWA A.;		
RL	INT. J. CANCER 58:602-607(1994).		
RP	[2]		
RN	SEQUENCE OF 34-354 FROM N.A.		
RX	MEDLINE: 94114699.		
RA	OTUDA M., UMEIDA A., MATSUMOTO Y., MOMOI Y., WATARI T., GOITSUKA R.,		
RA	O'BRIEN S.J., TSUJIMOTO H., HASEGAWA A.;		
RL	J. VET. MED. SCI. 55:801-805(1993).		
CC	-1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES		
CC	GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL		
CC	CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED		
CC	IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATOR, IT IS A		
CC	TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION		
CC	BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF		
CC	THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.		
CC	APOPTOSIS INDICATION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF		
CC	BAX AND BMS ANTISEN EXPRESSION, OR BY REPRESSION OF BCL-2		
CC	EXPRESSION.		
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.		
CC	-1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY		
CC	OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED		
CC	IN MANY TYPES OF CANCER.		
CC	-1- SIMILARITY: STRONG. TO P53 IN OTHER HIGHER EUKARYOTES.		
DR	EMBL: D26608; G538225; -;		
DR	EMBL: D16460; G575528; -;		
DR	PROSITE: PS00348; P53; 1.		
KM	ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;		
KM	NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.		
FT	DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).		
FT	DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).		
FT	MOD.RES 385 385 PHOSPHORYLATION (BY SIMILARITY).		

FT CONFLICT 285 285 K -> R (IN REF. 2).  
 SQ SEQUENCE 386 AA: 42692 MM: D6C7132A CRC32:  
 Query Match 100.0% Score 58; DB 1; Length 386;  
 Best Local Similarity 100.0% Pred. No. 1.48e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 354 GSRASHSHL 362  
 1 GSRASHSHL 9

RESULT 2  
 ID P53 CERAE STANDARD: PRT: 393 AA.  
 AC P13481:  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS CERCOPIITHECUS AETHIOPS (GREEN MONKEY) (GRIVET).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER.  
 RX MEDLINE; 90045967.  
 RA RIGAUDY P., ECKHARDT W.;  
 RL NUCLEIC ACIDS RES. 17:8375-8375(1989).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND BCL-2 EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC EMBL: X16384; G22796;  
 CC F1R; S06594; S06594.  
 CC DR HSSP; P04637; 10LG.  
 CC DR PROSITE; P500348; P53; 1.  
 CC KM ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 CC KM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 CC FT DOMAIN 1 68 ASP/GLU-RICH (ACIDIC).  
 CC FT DOMAIN 81 150 HYDROPHOBIC.  
 CC FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
 CC INTERACTION WITH DNA.  
 CC FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT MOD RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
 CC SQ SEQUENCE 393 AA: 43696 MM: BBEDDC62 CRC32:

Query Match 100.0% Score 58; DB 1; Length 393;  
 Best Local Similarity 100.0% Pred. No. 1.48e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRASHSHL 369  
 1 GSRASHSHL 9

RESULT 3  
 ID P53 HUMAN STANDARD: PRT: 393 AA.  
 AC P04637;  
 DT 13-AUG-1987 (REL. 05, CREATED)  
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
 GN TP53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85230577.  
 RA ZAKUT-HOORI R., BIENZ-TADMOR B., GIVOL D., OREN M.;  
 RL EMBO J. 4:1251-1255(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87064416.  
 RA LAMB P., CRAWFORD L.;  
 RL MOL. CELL. BIOL. 6:1379-1385(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85267676.  
 RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;  
 RL MOL. CELL. BIOL. 5:1601-1610(1985).  
 RN [4]  
 RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
 RX MEDLINE; 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RL ROTTER V.;  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89108008.  
 RA BICHMAN V.T., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P.,  
 RA GORGIYEV G.P.;  
 RL GENE 70:245-252(1988).  
 RN [6]  
 RP SEQUENCE OF 101-393 FROM N.A.  
 RX MEDLINE; 85126934.  
 RA MATLASHENSKI G., LAMB P., PIN D., PEACOCK J., CRAWFORD L.,  
 RA BENICIMOL S.;  
 RL EMBO J. 3:3257-3262(1984).  
 RN [7]  
 RP NUCLEAR LOCALIZATION SIGNAL.  
 RX MEDLINE; 90191730.  
 RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
 RL ONCOGENE 5:423-426(1990).  
 RN [8]  
 RP STRUCTURE BY NMR OF 319-360.  
 RX MEDLINE; 94294808.  
 RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
 RA APPELLA E., GROENBORN A.M.;  
 RL SCIENCE 265:386-391(1994).  
 RN [9]  
 RP STRUCTURE BY NMR OF 325-355.  
 RX MEDLINE; 95292092.  
 RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
 RL NAT. STRUCT. BIOL. 1:877-890(1994).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
 RX MEDLINE; 94294806.  
 RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
 RL SCIENCE 265:346-355(1994).  
 RN [11]  
 RP REVIEW.  
 RX MEDLINE; 94090335.  
 RA HARRIS C.C.;  
 RL SCIENCE 262:1980-1981(1993).  
 RN [12]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE; 91289156.  
 RA HOULSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
 RL SCIENCE 253:49-53(1991).  
 RN [13]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE; 96271983.  
 RA DE VRIES E.M.G., RICKLE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,



RA LIAO D., SOUSSI T., KOVACH J.S., SOMMER S.S.;  
 RL HUM. MUTAT. 7:202-213(1996).  
 RN [14]  
 RP VARIANT ARG-72.  
 RX MEDLINE: 91153807.  
 RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
 RL HUM. GENET. 86:369-370(1991).  
 RN [15]  
 RP VARIANT LFS THR-133.  
 RX MEDLINE: 92034774.  
 RA LAM J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
 RL CANCER RES. 51:6385-6387(1991).  
 RN [16]  
 RP VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
 RX MEDLINE: 91057657.  
 RA MARKIN D., LI F.P., STRONG L.C., FRAGMENT J.F. JR., NELSON C.E.,  
 RA KIM D.H., KASSEL J., GRKA M.A., BISCHOFF F.Z., TAINSKY M.A.,  
 RA FRIEND S.H.;  
 RL SCIENCE 250:1233-1238(1990).  
 RN [17]  
 RP VARIANT LFS ASP-245.  
 RX MEDLINE: 91080929.  
 RA SRIVASTAVA S., ZOU Z., PIROLO K., BLATTNER W., CHANG E.H.;  
 RL NATURE 348:747-749(1990).  
 RN [18]  
 RP VARIANT LFS LEU-272.  
 RX MEDLINE: 92147883.  
 RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
 RA POPLACK D.G., REMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
 RA KNUTSEN T., MINNA J.D.;  
 RL J. CLIN. INVEST. 89:640-647(1992).  
 RN [19]  
 RP VARIANTS LFS HIS-273 AND VAL-325.  
 RX MEDLINE: 92228023.  
 RA MARKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
 RA GEHARDT M.C., ANDERSEN T.I., BORESEN A.-L., LI F.P., GARBER J.,  
 RA STRONG L.C.;  
 RL NEW ENGL. J. MED. 326:1309-1315(1992).  
 RN [20]  
 RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
 RX MEDLINE: 90295284.  
 RA BARTER J., IGO R., GANNON J., LANE D.P.;  
 RL ONCOGENE 5:893-899(1990).  
 RN [21]  
 RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
 RX MEDLINE: 91017544.  
 RA RODRIGUES N.R., ROMAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
 RA GANNON J.V., LANE D.P.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 87:7555-7559(1990).  
 RN [22]  
 RP VARIANTS COLORECTAL CANCER MUTATIONS.  
 RX MEDLINE: 91282784.  
 RA ISHIOKA C., SATO T., GAMOH M., SUZUKI T., SHIBATA H., KANAMARU R.,  
 RA WAKUI A., YAMAZAKI T.;  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 177:901-906(1991).  
 RN [23]  
 RP VARIANTS OESOPHAGUS TUMORS L-152; A-155; H-175; F-176 AND H-273.  
 RX MEDLINE: 91330175.  
 RA CASSON A.G., MURHOPADHYAY T., CLEARY K.R., RO J.Y., LEVIN B.,  
 RA ROTH J.A.;  
 RL CANCER RES. 51:4495-4499(1991).  
 RN [24]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN CHINA.  
 RX MEDLINE: 91187113.  
 RA HSU I.C., METCALF R.A., SUN T., WELSH J.A., WANG N.J., HARRIS C.C.;  
 RL NATURE 350:427-428(1991).  
 RN [25]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN SOUTH AFRICA.  
 RX MEDLINE: 91182114.  
 RA BRESSAC B., KEM M., WANDS J., OZTURK M.;  
 RL NATURE 350:429-431(1991).  
 RN [26]  
 RP VARIANTS IN ANOGENITAL CARCINOMAS.

RX MEDLINE: 93010989.  
 RA CROOK T., VOIGDEN K.H.;  
 RL EMO J. 11:3935-3940(1992).  
 RN [27]  
 RP VARIANT WITH DUPLICATION OF PRO-177-HIS-178.  
 RX MEDLINE: 93265016.  
 RA BHATIA K., GUTIERREZ M.I., MAGRATH I.T.;  
 RL HUM. MOL. GENET. 1:207-208(1992).  
 RN [28]  
 RP VARIANTS IN BURKITT'S LYMPHOMAS.  
 RX MEDLINE: 93064692.  
 RA DUTHU A., DEBUIRE B., ROMANO J.W., EHRLHART J.C., FISCELLA M., MAY E.,  
 RA APPELLA E., MAY P.;  
 RL ONCOGENE 7:2161-2167(1992).  
 RN [29]  
 RP VARIANT NASOPHARYNGEAL CARCINOMA THR-280.  
 RX MEDLINE: 92335329.  
 RA SUN Y., HEGAMAYER G., HENG Y.-J., HILDESHEIM A., CHEN J.-Y., CAO Y.,  
 RA YAO K.-T., COLBURN N.H.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:6516-6520(1992).  
 RN [30]  
 RP VARIANTS IN COLON TUMORS.  
 RX MEDLINE: 9330562.  
 RA HAMELIN R., JEGO N., LAURENT-PUIG P., VIDAUD M., THOMAS G.;  
 RL ONCOGENE 8:2213-2220(1993).  
 RN [31]  
 RP CHARACTERIZATION OF VARIANT ALA-143.  
 RX MEDLINE: 9428378.  
 RA ZHANG W., GUO X.-Y., HU G.-Y., LIU W.-B., SHAY J.W., DEISSEROTH A.B.;  
 RL EMO J. 13:2535-2544(1994).  
 RN [32]  
 RP VARIANTS LFS HIS-175; ARG-193; GLN-248; CYS-273 AND TYR-275.  
 RX MEDLINE: 95193787.  
 RA FREBOURG T., BARBIER N., YAN Y.-X., GARBER J.E., DREYFUS M.,  
 RA FRAMENI J.F. JR., LI F.P., FRIEND S.H.;  
 RL AM. J. HUM. GENET. 56:608-615(1995).  
 RN [33]  
 RP VARIANT LFS HIS-175.  
 RX MEDLINE: 96423319.  
 RA VARLEY J.M., MCGROWN G., THORNCROFT M., TRICKER K.J., TEARE M.D.,  
 RA SANTIBANEZ-KOREF M.F., HOLLSTON R.S., MARTIN J., BIRCH J.M.,  
 RA EVANS D.G.R.;  
 RL J. MED. GENET. 32:942-945(1995).  
 RN [34]  
 RP VARIANTS BA PHE-176; SER-245; TRP-248; TRP-282 AND GLN-286.  
 RX MEDLINE: 96233927.  
 RA AUDREZET M.-P., ROBASZKIEWICZ M., MERCIER B., NOUSBAUM J.-B.,  
 RA HARDY E., BAIL J.-P., VOLANT A., LOZAC'H P., GOUEROU H., FEREC C.;  
 RL HUM. MUTAT. 7:109-113(1996).  
 RN

\*\*\*: remainder of annotations omitted.

Query Match 100.0%; Score 58; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1,48e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB	361 GSRASHSHL 369	1 GSRASHSHL 9
QY		
RESULT	4	
ID	P53-SHEEP	STANDARD; PRT; 382 AA.
AC	P51664;	
DT	01-OCT-1996 (REL. 34, CREATED)	
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)	
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)	
DE	CELLULAR TUMOR ANTIGEN P53.	
GN	P53.	
OS	OVIS ARIES (SHEEP).	
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;	
CC	EUTHERIA; ARTIODACTYLA.	
RN	[1]	

RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD:  
RX DEQUEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
RL DNA SEQ. 5:255-259(1995).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC EMBL: X81705; G602357; -  
DR PROSITE: PS00348; P53; 1.  
KM ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KM NUCLEAR PROTEIN: PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 1 66 ASP/GLU-RICH (ACIDIC).  
FT MOD.RES 300 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD.RES 381 381 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 382 AA; 42809 MW; 0CB99A00 CRC32;  
Query Match 87.9%; Score 51; DB 1; Length 382;  
Best Local Similarity 100.0%; Pred. No. 1,73e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 351 SRAHSHL 358  
QY 2 SRAHSHL 9  
RESULT 5  
ID P53 BOVIN STANDARD; PRT; 386 AA.  
AC 029628;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
DE TP53.  
OS BOS TAURUS (BOVINE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; ANTIDACTYLIA.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE-LIVER;  
RX MEDLINE: 95352829.  
RA DEQUEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
RL DNA SEQ. 5:261-264(1995).  
RN (2)  
RP SEQUENCE OF 13-386 FROM N.A.  
RC STRAIN-HOLSTEIN; TISSUE-THYMUS;  
RA MEDLINE: 96401400.  
RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;  
RL VET. IMMUNOL. IMMUNOPATHOL. 52:53-63(1996).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

DR EMBL: X81704; G602333; -  
DR PROSITE: DA9825; G1729419; -  
DR ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KM NUCLEAR PROTEIN: PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
FT MOD.RES 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD.RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD.RES 380 380 R -> T (IN REF. 2).  
SQ SEQUENCE 386 AA; 43255 MW; 0322BF3D CRC32;  
Query Match 87.9%; Score 51; DB 1; Length 386;  
Best Local Similarity 100.0%; Pred. No. 1,73e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 355 SRAHSHL 362  
QY 2 SRAHSHL 9  
RESULT 6  
ID P53-RABBIT STANDARD; PRT; 391 AA.  
AC 095330;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
DE TP53.  
OS ORYCTOLAGUS CUNICULUS (RABBIT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; LAGOMORPHA.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-NEW ZEALAND;  
RX LE GOAS F., MAY P., RONCO P., CARON DE FROMENTEL C.;  
RL GENE 185:169-173(1997).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC EMBL: X90592; E194962; -  
DR PROSITE: PS00348; P53; 1.  
KM ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KM NUCLEAR PROTEIN: PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 1 70 ASP/GLU-RICH (ACIDIC).  
FT MOD.RES 308 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD.RES 390 390 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 391 AA; 43435 MW; 30A36172 CRC32;  
Query Match 84.5%; Score 49; DB 1; Length 391;  
Best Local Similarity 88.9%; Pred. No. 6,33e-02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 359 GSRASHL 367  
QY 1 GSRASHL 9  
RESULT 7  
ID A1AH HUWAN STANDARD; PRT; 201 AA.  
AC P19652;

DT 01-FEB-1991 (REL. 17, CREATED)  
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ALPHA-1-ACID GLYCOPROTEIN 2 PRECURSOR (AGP 2) (OROSOMUCOID 2) (OMD 2).  
 GN ORM2 OR AGP2.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUETHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 88029318.  
 RA DENTE L., PIZZA M.G., METSPALU A., CORTESE R.;  
 RN EMO J. 6:2289-2296(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 88329732.  
 RA MERRITT C.M., BOARD P.G.;  
 RN GENE 66:97-106(1988).  
 RN [3]  
 RP DISULFIDE BONDS.  
 RX MEDLINE; 74290014.  
 RA SCHMID K., BOERGT W., COLLINS J.H., NANNO S.;  
 RN BIOCHEMISTRY 13:2694-2697(1974).  
 RN [4]  
 RP CARBOHYDRATE-BINDING SITES.  
 RX MEDLINE; 92231810.  
 RA TREUHEIT M.J., COSTELLO C.E., HALLSALL H.B.;  
 RN BIOCHEM. J. 283:105-112(1992).  
 CC -1- FUNCTION: APPEARS TO FUNCTION IN MODULATING THE ACTIVITY OF THE  
 CC IMMUNE SYSTEM DURING THE ACUTE-PHASE REACTION.  
 CC -1- INDUCTION: ALPHA-1-AGP IS SYNTHESIZED IN THE LIVER, THE  
 CC SYNTHESIS BEING CONTROLLED BY GLUCOCORTICOID, INTERLEUKIN-1  
 CC AND INTERLEUKIN-6, IT INCREASES 5- TO 50-FOLD UPON INFLAMMATION.  
 CC -1- POLYMORPHISM: MANY DIFFERENT VARIANTS OF ORM2 ARE KNOWN.  
 CC -1- SIMILARITY: THIS PROTEIN BELONGS TO THE FAMILY OF SMALL  
 CC HYDROPHOBIC MOLECULES TRANSPORT PROTEINS (LIPOCALINS).  
 CC EMBL; X06675; G757910; ALT\_SEQ.  
 CC EMBL; X06674; G757909; ALT\_SEQ.  
 CC EMBL; X06676; G24467; -  
 CC EMBL; X05780; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL; X06678; G24470; -  
 CC EMBL; X06679; G24472; -  
 CC EMBL; X06680; E34408; -  
 CC EMBL; X05784; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL; M21540; G177840; -  
 CC PIR; J70326; J70326.  
 CC PIR; B28346; B28346.  
 CC SWISS-2DPAGE; P19652; HUMAN.  
 CC MIM; 136610; -  
 CC PROSITE; PS00213; LIPOCALIN; 1.  
 CC KW GLYCOPROTEIN; PLASMA; ACUTE PHASE; SIGNAL; LIPOCALIN;  
 CC MULTIGENE FAMILY.  
 CC FT STGNL 1 18  
 CC FT CHAIN 19 201  
 CC FT MOD\_RES 19 19  
 CC FT DISULFID 23 165  
 CC FT DISULFID 90 183  
 CC FT CARBOHYD 33 33  
 CC FT CARBOHYD 56 56  
 CC FT CARBOHYD 72 72  
 CC FT CARBOHYD 93 93  
 CC FT CARBOHYD 103 103  
 CC FT CONFLICT 119 119  
 CC FT SEQUENCE 201 AA; 23602 MW; 105ED8CF CRC32;  
 CC L -> N (IN REF. 1).  
 Query Match 74.1%; Score 43; DB 1; Length 201;  
 Best Local Similarity 55.6%; Pred. No. 2,53e+00;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 8  
 ID NTLM STRPU STANDARD; PRT; 97 AA.  
 AC P15354;  
 DT 01-APR-1990 (REL. 14, CREATED)  
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4L (EC 1.6.5.3).  
 GN ND4L.  
 OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).  
 OG MITOCHONDRION.  
 OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;  
 OC EUECHINOIDEA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89011951.  
 RA JACOBS H.T., ELLIOTT D.J., MATH V.B., FARQUHARSON A.;  
 RN J. MOL. BIOL. 202:185-217(1988).  
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.  
 CC EMBL; X12631; E74291; ALT\_SEQ.  
 CC PIR; S01502; S01502.  
 CC KX OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.  
 CC KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.  
 CC SEQUENCE 97 AA; 10610 MW; 74E4F9B4 CRC32;  
 Query Match 72.4%; Score 42; DB 1; Length 97;  
 Best Local Similarity 75.0%; Pred. No. 4,54e+00;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 81 SRTSSNL 88  
 QY 2 SRAHSHL 9  
 RESULT 9  
 ID SR21\_YEAST STANDARD; PRT; 166 AA.  
 AC P32342;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
 DE SIGNAL RECOGNITION PARTICLE 21 KD PROTEIN (SRP21).  
 GN SRP21 OR YKL122C OR YKL527.  
 OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92383952.  
 RA COLLEAUX L., RICHARD G.-F., THIERRY A., DUJON B.;  
 RN YEAST 8:325-336(1992).  
 RN [2]  
 RP SEQUENCE OF 1-14; 44-54; 74-84 AND 106-119, AND ACETYLATION.  
 RC STRAIN-S288C;  
 RX MEDLINE; 95009940.  
 RA BROWN J.D., HANN B.C., MEDZHIRADSKY K.F., NIMA M., BURLINGAME A.L.,  
 RA WALTER P.;  
 RN EMO J. 13:4390-4400(1994).  
 CC -1- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE  
 CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC  
 CC RETICULUM MEMBRANE. IT MUST BE INVOLVED INTIMATELY IN THE  
 CC TRANSLOCATION OF A WIDE VARIETY OF PROTEIN SUBSTRATES. SRP21  
 CC COULD POSSIBLY BIND TO SCR1.  
 CC -1- SUBUNIT: YEAST SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA  
 CC MOLECULE (SCR1) AND AT LEAST SEVEN PROTEIN SUBUNITS: SRP72, SRP68,  
 CC SRP54, SEC65, SRP21, SRP14 AND SRP7.  
 CC EMBL; S44213; G254451; -  
 CC DR EMBL; Z28122; G486206; -  
 CC DR PIR; S25360; S25360.  
 CC DR SGD; L0002062; SRP21.  
 CC KW SIGNAL RECOGNITION PARTICLE; ACETYLATION; RNA-BINDING.  
 CC FT INIT\_MET 0 0  
 CC FT MOD\_RES 1 1  
 CC FT DOMAIN 154 166  
 CC FT SEQUENCE 166 AA; 18294 MW; 6978BA92 CRC32;  
 CC ACETYLATION.  
 CC LYS-RICH (HIGHLY BASIC).

Query Match	72.4%;	Score 42;	DB 1;	Length 166;
Best Local Similarity	71.4%;	Pred. No. 4.54e+00;		
Matches	5;	Conservative	2;	Mismatches 0;
				Indels 0;
				Gaps 0;

Db	43	RTHNSHL	49
QY	3	RAHSSHL	9

RESULT	10			
ID	YPBG_BACSU	STANDARD;	PRT;	259 AA.
10	YPBG_BACSU			

OS BACILLUS SUBTILIS.  
OC PROKARYOTA; FIRINICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE  
RN [1]  
RP SEQUENCE FROM N.A.

Query Match	72.48;	Score 42;	DB 1;	Length 259;
Best Local Similarity	50.08;	Pred. No. 4.54e+00;		
Matches	4;	Conservative	4;	Mismatches 0; Indels 0; Gaps

QY : 2 SRAHSHL 9

RESULT	11
ID YX17.MYCTU	STANDARD; PRT; 290 AA

[illegible]

```
Query Match      72.48; Score 42; DB 1; Length 290;
Best Local Similarity 55.68; Pred. No. 4.54e+00;
Matches      5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
Db 115 GPRGHAYHL 123
    1:1:1:1:1
QY 1 GSRHSSHL 9
```

## RESULT 12

```

ID      P53 MOUSE          STANDARD:          PR1:    390 AA.
DT      P02340:
DT      21-JUL-1986 (REL. 01, CREATED)
DT      01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT      01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE      CELLULAR TUMOR ANTIGEN P53.
GN      TP53 OR TRP53 OR P53.
OS      MUS MUSCULUS (MOUSE).
OC      EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:
CC      [1]
CC      RA      SEQUENCE FROM N.A.
CC      RX      MEDLINE: 85027173.
CC      RL      BIENZ B., ZAKUT-HOORI R., GIVOL D., OREN M.:
CC      RN      EMBO J. 3:2179-2183(1984).
CC      [2]
CC      RA      SEQUENCE FROM N.A.
CC      RX      MEDLINE: 84068204.
CC      RL      ZAKUT-HOORI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIVOL D.:
CC      RN      NATURE 306:594-597(1983).
CC      [3]
CC      RA      SEQUENCE FROM N.A.
CC      RX      MEDLINE: 84272240.
CC      RL      JENKINS J.R., RUDGE K., REDMOND S., MADE-EVANS A.:
CC      RN      NUCLEIC ACIDS RES. 12:5609-5626(1984).
CC      [4]
CC      RA      SEQUENCE FROM N.A. (CLONES PC053; P53-M11 AND P53-M8).
CC      RX      MEDLINE: 87064640.
CC      RL      ARII N., NOKURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.,
CC      RN      ROTTER V.:
CC      [5]
CC      RA      MOL. CELL. BIOL. 6:3232-3233(1986).
CC      RN      [5]
CC      RA      SEQUENCE OF 222-258 FROM N.A.
CC      RX      MEDLINE: 92115342.
CC      RL      BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BREMNER R.,
CC      RN      BALMAIN A.:
CC      [6]
CC      RA      ONCOGENE 6:2363-2369(1991).
CC      RN      [6]
CC      RA      PHOSPHORYLATION SITES.
CC      RX      MEDLINE: 86149247.
CC      RL      SAMAD A., ANDERSON C.W., CARROLL R.B.:
CC      RN      PROC. NATL. ACAD. SCI. U.S.A. 83:897-901(1986).
CC      [7]
CC      RA      PHOSPHORYLATION SITES.
CC      RX      MEDLINE: 91006019.
CC      RL      MEER D.W., SIMON S., KIKKAWA U., ECKHART W.:
CC      RN      EMBO J. 9:3253-3260(1990).
CC      [8]
CC      RA      FUNCTION: ACT AS A TUMOR SUPPRESSOR. IN MANY TUMOR TYPES, INDUCES
CC      RX      GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC      RL      CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN
CC      RN      TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC      [9]
CC      RA      TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC      RX      BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC      RL      THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC      RN      APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC      [10]
CC      RA      BAX AND BAX ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC      RX      EXPRESSION.
CC      RL      [11]
CC      RA      SUBCELLULAR LOCATION: NUCLEAR.
CC      RX      [12]
CC      RA      DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC      RX      OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC      RL      IN MANY TYPES OF CANCER.
CC      [13]
CC      RA      -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.
CC      RX      DR      EMBL: X00876: 6871421: -.
CC      RL      DR      EMBL: X00877: 6871421: JOINED.
CC      RN      DR      EMBL: X00878: 6871421: JOINED.
CC      RX      DR      EMBL: X00879: 6871421: JOINED.
CC      RL      DR      EMBL: X00880: 6871421: JOINED.
CC      RN      DR      EMBL: X00881: 6871421: JOINED.
CC      RX      DR      EMBL: X00882: 6871421: JOINED.
CC      RL      DR      EMBL: X00883: 6871421: JOINED.
CC      RN      DR      EMBL: X00884: 6871421: JOINED.
CC      RX      DR      EMBL: X00885: 6871421: JOINED.
CC      RL      DR      EMBL: X01700: 6200205: -.

```

FT	DOMAIN	34	559		EXTRACELLULAR (POTENTIAL).
FT	TRANSFM	560	580		POTENTIAL.
FT	DOAIN	581	998		CYTOPLASMIC (POTENTIAL).
FT	NP_BIND	633	896		PROTEIN KINASE.
FT	BINDING	639	647		ATP (BY SIMILARITY).
FT	ACT_SITE	665	665		ATP (BY SIMILARITY).
FT	CARBOHYD	758	758		BY SIMILARITY.
FT	CARBOHYD	351	351		POTENTIAL.
FT	CARBOHYD	445	445		POTENTIAL.
SQ	SEQUENCE	998 AA;	110286 MW;	6FB75A43 CRC32;	
Query Match					
	Best Local Similarity	72.4%;	Score 42;	DB 1;	Length 998;
	Matches	5; Conservative	Pred. NO. 4.34e+00;	Mismatches 2;	Indels 0; Gaps 0
Dd	408 PRVHTSHL 415				
Oy	: ::				
	2 SRAHSSH 9				
RESULT 14					
ID	WNT1_XENLA	STANDARD;	PRT;	371 AA.	
AC	P10108:				
DT	01-MAR-1989 (REL. 10, CREATED)				
DT	01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)				
DT	01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)				
DE	XWNT-1 PROTEIN PRECURSOR (XINT-1).				
GN	WNT-1 OR INT-1.				
OS	XENOPUS LAEVIS (AFRICAN CLAWED FROG).				
OC	EUMARIOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 89098373.				
RA	NORMDENNER J., MEJLINK F., VERRIJER P., RIJSEWIJK F., DESTREE O.;				
RL	NUCLEIC ACIDS RES. 17:11-18(1989).				
RN	[2]				
RP	FUNCTION.				
RX	MEDLINE; 89376559.				
RA	MCMAHON A.P., MOON R.T.;				
RL	CELL 58:1075-1084(1989).				
CC	-1- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING				
CC	MOLECULE IMPORTANT IN CNS DEVELOPMENT. IS LIKELY TO SIGNAL OVER				
CC	ONLY FEW CELL DIAMETERS.				
CC	-1- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE				
CC	EXTRACELLULAR MATRIX.				
CC	-1- TISSUE SPECIFICITY: AT NEURONA IN ANTERIOR NEURAL FOLD; AT TAILBUD				
CC	IN DORSAL MIDLINE OF MIDLRAIN.				
CC	-1- DEVELOPMENTAL STAGE: NEURULA ONWARDS.				
CC	-1- SIMILARITY: TO OTHER MEMBERS OF THE WNT FAMILY.				
DR	EMBL; X13138; 665236; -.				
DR	PIR; S02113; TYXL1.				
DR	PROSITE; PS00246; WNT1; 1.				
RW	DEVELOPMENTAL PROTEIN; GLYCOPROTEIN; SIGNAL.				
FT	SIGNAL	1	19		POTENTIAL.
FT	CHAIN	20	371		XWNT-1 PROTEIN.
FT	CARBOHYD	28	28		POTENTIAL.
FT	CARBOHYD	261	261		POTENTIAL.
FT	CARBOHYD	317	317		POTENTIAL.
FT	CARBOHYD	360	360		POTENTIAL.
SQ	SEQUENCE	371 AA;	41125 MW;	AB9CE866 CRC32;	
Query Match					
	Best Local Similarity	70.7%;	Score 41;	DB 1;	Length 371;
	Matches	5; Conservative	Pred. No. 8.03e+00;	Mismatches 3;	Indels 0; Gaps 0;
Dd	267 GSRSDDPHL 275				
Oy	::: ::				
	1 GSRAHSSH 9				
RESULT 15					
ID	YOR3_GIUSU	STANDARD;	PRT;	444 AA.	
AC	O05543;				

DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL PROTEIN IN ADHS 5' REGION (ORF3) (FRAGMENT).  
OS GLUCONOBACTER SUBOXIDANS.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC ACETOBACTERACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FO 12528;  
RX MEDLINE; 97208225.  
RA KONDO K., HORINOCHI S.;  
RL APPL. ENVIRON. MICROBIOL. 63:1131-1138(1997).  
DR EMBL; D86440; D1020536;  
KW HYPOTHETICAL PROTEIN.  
FT NON\_TER 1  
SO SEQUENCE 444 AA; 48225 MW; 52A074F4 CRC32;

Query Match 70.7%; Score 41; DB 1; Length 444;  
Best Local Similarity 55.6%; Pred. No. 8.03e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 45 GSRSHPRHY 53  
|||:|:|:  
OY 1 GSRASHSHL 9

Search completed: Fri Sep 11 13:57:11 1998  
Job time : 12 secs.

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697. 698. 699. 700. 701. 702. 703. 704. 705. 706. 707. 708. 709. 710. 711. 712. 713. 714. 715. 716. 717. 718. 719. 720. 721. 722. 723. 724. 725. 726. 727. 728. 729. 730. 731. 732. 733. 734. 735. 736. 737. 738. 739. 740. 741. 742. 743. 744. 745. 746. 747. 748. 749. 750. 751. 752. 753. 754. 755. 756. 757. 758. 759. 760. 761. 762. 763. 764. 765. 766. 767. 768. 769. 770. 771. 772. 773. 774. 775. 776. 777. 778. 779. 780. 781. 782. 783. 784. 785. 786. 787. 788. 789. 790. 791. 792. 793. 794. 795. 796. 797. 798. 799. 800. 801. 802. 803. 804. 805. 806. 807. 808. 809. 810. 811. 812. 813. 814. 815. 816. 817. 818. 819. 820. 821. 822. 823. 824. 825. 826. 827. 828. 829. 830. 831. 832. 833. 834. 835. 836. 837. 838. 839. 840.

Release 3.1A John F. Collins, Biocomputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Fri Sep 11 13:57:28 1998;  MasPar time 3.93 Seconds
            95 450 M111400 2001 updates/Sec
```

Tabular output not generated.

|              |                           |
|--------------|---------------------------|
| Title:       | >US-08-452-843-21         |
| Description: | (1-9) from US08452843.pdf |
| Parent Case: | FO                        |

Sequence: . 1 GSRAHSSL 9

Scoring table: PAM 150

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 0% Listing first 45 summaries

Database

1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mhc 6:sp\_organelle 7:sp\_phage 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp\_vertebrate  
13:sp\_unclassified

Statistics: Mean 19.763; Variance 19.325; scale 1.023

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description                         | Pred. No |
|------------|-------|-------------|--------|----|--------|-------------------------------------|----------|
| 1          | 58    | 100.0       | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN              | 1,82e+00 |
| 2          | 58    | 100.0       | 393    | 2  | Q16811 | CELLULAR TUMOR ANTIGEN              | 1,82e+00 |
| 3          | 58    | 100.0       | 393    | 2  | Q16807 | CELLULAR TUMOR ANTIGEN              | 1,82e+00 |
| 4          | 58    | 100.0       | 393    | 2  | Q16808 | CELLULAR TUMOR ANTIGEN              | 1,82e+00 |
| 5          | 58    | 100.0       | 393    | 2  | Q16848 | CELLULAR TUMOR ANTIGEN              | 1,82e+00 |
| 6          | 58    | 100.0       | 393    | 2  | Q16810 | CELLULAR TUMOR ANTIGEN              | 1,82e+00 |
| 7          | 58    | 100.0       | 393    | 2  | Q15085 | P53 TRANSFORMATION SUP              | 1,82e+00 |
| 8          | 58    | 100.0       | 393    | 2  | Q15635 | P53 TRANSFORMATION SUP              | 1,82e+00 |
| 9          | 58    | 100.0       | 393    | 2  | Q16809 | CELLULAR TUMOR ANTIGEN              | 1,82e+00 |
| 10         | 58    | 100.0       | 393    | 2  | Q15088 | P53 TRANSFORMATION SUP              | 1,82e+00 |
| 11         | 58    | 100.0       | 393    | 2  | Q15087 | P53 TRANSFORMATION SUP              | 1,82e+00 |
| 12         | 44    | 75.9        | 301    | 9  | Q53796 | BLEOMYCIN ACETYLTRANSFERASE         | 1,82e+00 |
| 13         | 43    | 74.1        | 201    | 2  | Q16571 | ALPHA-1-ACID GLYCOPROTEIN           | 1,73e+00 |
| 14         | 43    | 72.4        | 267    | 4  | Q29495 | ARYLAALKYLAMINE N-ACETYLTRANSFERASE | 3,14e+00 |
| 15         | 42    | 72.4        | 286    | 11 | P89003 | P53 (FRAGMENT)                      | 5,65e+00 |
| 16         | 42    | 72.4        | 378    | 11 | P90332 | P53 (FRAGMENT)                      | 5,65e+00 |
| 17         | 42    | 72.4        | 378    | 11 | P89002 | GLYCOPROTEIN HOMOLOGUE              | 5,65e+00 |
| 18         | 42    | 72.4        | 384    | 11 | Q88523 | KLANKO150 PROTEIN (FRAG             | 5,65e+00 |
| 19         | 42    | 72.4        | 944    | 2  | Q14163 | KLANKO150 PROTEIN (FRAG             | 5,65e+00 |
| 20         | 42    | 72.4        | 3268   | 1  | Q03280 | DB035.LP.                           | 5,65e+00 |

|    |    |      |      |    |        |                         |          |
|----|----|------|------|----|--------|-------------------------|----------|
| 21 | 41 | 70.7 | 93   | 9  | Q27549 | PUSM GENE ENCODING 5'-  | 1.01e+01 |
| 22 | 41 | 70.7 | 360  | 3  | Q21495 | COGMD MO3D4.            | 1.01e+01 |
| 23 | 41 | 70.7 | 400  | 9  | Q49949 | U1756C.                 | 1.01e+01 |
| 24 | 41 | 70.7 | 406  | 11 | Q11374 | HYPOTHELTICAL 44.0 KD P | 1.01e+01 |
| 25 | 41 | 70.7 | 473  | 8  | Q10353 | 1-MINOCYCLOPROPANE-1-   | 1.01e+01 |
| 26 | 41 | 70.7 | 533  | 10 | Q53240 | PROTON GATED CATION CH  | 1.01e+01 |
| 27 | 41 | 70.7 | 642  | 11 | Q98310 | MC14AR.                 | 1.01e+01 |
| 28 | 41 | 70.7 | 1534 | 4  | Q28298 | RIBOSOME RECEPTOR.      | 1.01e+01 |
| 29 | 40 | 68.0 | 92   | 9  | Q03139 | 741E.                   | 1.77e+01 |
| 30 | 40 | 68.0 | 151  | 10 | Q63119 | MRNA FOR C-MET PROTO-O  | 1.77e+01 |
| 31 | 40 | 68.0 | 346  | 6  | Q33240 | MAURASE (FRAGMENT).     | 1.77e+01 |
| 32 | 40 | 68.0 | 523  | 3  | Q18236 | SIMILAR TO HISTIDINE A  | 1.77e+01 |
| 33 | 40 | 68.0 | 533  | 9  | Q53903 | 6 ACTVA REGION GENES O  | 1.77e+01 |
| 34 | 40 | 68.0 | 598  | 2  | Q14181 | DNA POLYMERASE ALPHA.   | 1.77e+01 |
| 35 | 40 | 68.0 | 959  | 1  | Q06337 | CHROMOSOME IV COSMID 9  | 1.77e+01 |
| 36 | 40 | 68.0 | 1382 | 10 | P97523 | HGT RECEPTOR PRECURSOR  | 1.77e+01 |
| 37 | 40 | 68.0 | 1382 | 10 | P97579 | HEPATOCYTE GROWTH FACT  | 1.77e+01 |
| 38 | 40 | 68.0 | 1807 | 10 | Q64632 | INTEGRIN BETA 4 SUBUNIT | 1.77e+01 |
| 39 | 39 | 67.2 | 315  | 12 | P79930 | XENOPUS ATRONAL HOMOLOC | 3.09e+01 |
| 40 | 39 | 67.2 | 364  | 11 | Q68795 | N5S, ISOLATE JK030 (FR  | 3.09e+01 |
| 41 | 39 | 67.2 | 419  | 9  | Q35415 | YRKM2.                  | 3.09e+01 |
| 42 | 39 | 67.2 | 706  | 11 | Q39286 | COUNTERPART OF HSV-1 G  | 3.09e+01 |
| 43 | 39 | 67.2 | 3019 | 11 | Q68801 | POLYPROTEIN.            | 3.09e+01 |
| 44 | 39 | 67.2 | 3493 | 11 | Q39734 | POLYPROTEIN.            | 3.09e+01 |
| 45 | 39 | 67.2 | 5262 | 2  | Q14686 | ALR.                    | 3.09e+01 |

## ALIGNMENTS

|        |   |              |                     |               |             |
|--------|---|--------------|---------------------|---------------|-------------|
| ID     | 029475  | 1            | PRELIMINARY;        | PRT;          | 261 AA.     |
| AC     | Q29475:   |              |                     |               |             |
| DT     | 01-NOV-1996 (TREMBLREL. 01,                                       |              |                     |               |             |
| DI     | 01-NOV-1996 (TREMBLREL. 01,                                       |              |                     |               |             |
| DE     | 01-JAN-1998 (TREMBLREL. 05,                                       |              |                     |               |             |
| DD     | CELLULAR TUMOR ANTIGEN P53  |              |                     |               |             |
| DS     | P53.  |              |                     |               |             |
| GN     | CANIS FAMILIARIS (DOG).   |              |                     |               |             |
| OC     | EUDAROTA; METAFOA; CHORDATA;                                      |              |                     |               |             |
| CC     | EUTHERIA; CANINIOVA.  |              |                     |               |             |
| RN     | [1]   |              |                     |               |             |
| RP     | SEQUENCE FROM N.A.  |              |                     |               |             |
| RC     | TISSUE-MAMMARY GLAND;   |              |                     |               |             |
| RX     | MEDLINE; 97194812.  |              |                     |               |             |
| RA     | VAN LIEUWEN I., RUTTEMAN G.R.,                                    |              |                     |               |             |
| RL     | DEVILJE P.; HELLMEN E., CORNELISSE C.C.J.,                        |              |                     |               |             |
| RA     | ANTICANCER RES. 16:3737-3744(1996).                               |              |                     |               |             |
| CC     | -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT |              |                     |               |             |
| CC     | PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL   |              |                     |               |             |
| CC     | CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY |              |                     |               |             |
| CC     | REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED |              |                     |               |             |
| CC     | FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF   |              |                     |               |             |
| CC     | CYCLIN-DEPENDENT KINASES (BY SIMILARITY).                         |              |                     |               |             |
| DR     | -1- SUBCELLULAR LOCATION: NUCLEAR.                                |              |                     |               |             |
| DR     | EMBL; L37107; G1463021..  |              |                     |               |             |
| KM     | PROSITE; PS00346; P53; 1.   |              |                     |               |             |
| KW     | ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  |              |                     |               |             |
| KW     | NUCLEAR PROTEIN; PHOSPHORYLATION.                                 |              |                     |               |             |
| FT     | NON_TER 1   |              |                     |               |             |
| FT     | NON_TER 1   |              |                     |               |             |
| SO     | SEQUENCE 281 AA; 31762 MW; FC7BAE31 CRC32;                        |              |                     |               |             |
| Db     | Query Match   | 100.0%;      | Score 58;           | DB 4;         | Length 281; |
| Oy     | Best Local Similarity 100.0%;                                     |              | Pred. NO. 1.82e-04; |               |             |
|        | Matches 9;  | Conservative | 0;                  | Mismatches 0; | Indels 0;   |
|        | Gaps 0;   |              |                     |               |             |
| Db     | 254 GSRAHSHTL 262   |              |                     |               |             |
| Oy     | 1 GSRAHSHTL 9   |              |                     |               |             |
| RESULT | 2   |              |                     |               |             |
| ID     | Q16811  |              | PRELIMINARY;        | PRT;          | 393 AA.     |

AC 016811: CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
 CC EUTHERIA: PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE: 85126934.  
 RA MATTAS, K. S.; LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
 RA BENCHIMOL S.;  
 RL EMO J. 10:2879-2887(1991).  
 RL EMO J. 10:2879-2887(1991).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87064416.  
 RA LAMB P., CRAWFORD L.;  
 RL MOL. CELL. BIOL. 6:1379-1385(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: M13121; G386994; JOINED.  
 DR EMBL: M13112; G386994; JOINED.  
 DR EMBL: M13113; G386994; JOINED.  
 DR EMBL: M13114; G386994; JOINED.  
 DR EMBL: M13115; G386994; JOINED.  
 DR EMBL: M13116; G386994; JOINED.  
 DR EMBL: M13117; G386994; JOINED.  
 DR EMBL: M13118; G386994; JOINED.  
 DR EMBL: M13119; G386994; JOINED.  
 DR EMBL: M13120; G386994; JOINED.  
 DR PROSITE: PS00348; P53; 1.  
 CC REPEAT: TUMOR ANTIGEN; ANTI-ONCOGENE; DNA-BINDING;  
 CC TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN;  
 CC PHOSPHORYLATION.  
 KW NON\_TER 393  
 FT SEQUENCE 393 AA; 43698 MW; 3EA71A31 CRC32;  
 SO  
 Query Match 100.0%; Score 58; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1.82e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 361 GSRASHSHL 369  
 Oy 1 GSRASHSHL 9  
 RESULT 3  
 ID 016807: PRELIMINARY; PRT; 393 AA.  
 AC 016807:  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
 CC EUTHERIA: PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., SHANAHAN F., YOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: X60011; G506435; -.  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 193 193 R->H.  
 FT NON\_TER 393 393  
 SO SEQUENCE 393 AA; 43731 MW; 279BC9CB CRC32;  
 Query Match 100.0%; Score 58; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1.82e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 361 GSRASHSHL 369  
 Oy 1 GSRASHSHL 9  
 RESULT 4  
 ID 016808: PRELIMINARY; PRT; 393 AA.  
 AC 016808:  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
 CC EUTHERIA: PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., SHANAHAN F., YOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: X60018; G506449; -.  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 163 163 H->Y.  
 FT NON\_TER 393 393  
 SO SEQUENCE 393 AA; 43627 MW; AFD8A9E3 CRC32;  
 Query Match 100.0%; Score 58; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1.82e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 361 GSRASHSHL 369  
 Oy 1 GSRASHSHL 9  
 RESULT 5  
 ID 016848: PRELIMINARY; PRT; 393 AA.  
 AC 016848:  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
 CC EUTHERIA: PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87089826.



RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTER V.,  
 RL MOL. CELL. BIOL. 6:4650-4656(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: M14694: G339814; -  
 DR POSTITE: P500348; P53; 1.  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ACTIVATOR.  
 KM TRANSCRIPTION REGULATION; ACTIVATOR.  
 SO SEQUENCE 393 AA: 43723 MW; DA7D302F CRC32;  
 Query Match 100.0%; Score 58; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1.82e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 361 GSRHSSHL 369  
 OY 1 GSRHSSHL 9  
 RESULT 6  
 ID 016810 PRELIMINARY; PRT; 393 AA.  
 AC 016810;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: X60020: G506453; -  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 254 254 D -> N.  
 FT VARIANT 254 254 D -> V.  
 FT NON-TER 393 393  
 SO SEQUENCE 393 AA: 43714 MW; 5F914579 CRC32;  
 Query Match 100.0%; Score 58; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1.82e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 361 GSRHSSHL 369  
 OY 1 GSRHSSHL 9  
 RESULT 7  
 ID 015086 PRELIMINARY; PRT; 393 AA.  
 AC 015086;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBO J. 10:2879-2887(1991).  
 DR EMBL: X60013: G506439; -  
 DR VARIANT 246 246 T -> M.  
 FT NON-TER 393 393  
 SO SEQUENCE 393 AA: 43682 MW; 943B62A3 CRC32;  
 Query Match 100.0%; Score 58; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1.82e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 361 GSRHSSHL 369  
 OY 1 GSRHSSHL 9  
 RESULT 8  
 ID 016535 PRELIMINARY; PRT; 393 AA.  
 AC 016535;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBO J. 10:2879-2887(1991).  
 DR EMBL: X60017: G506447; -  
 DR EMBL: X60015: G506443; -  
 FT VARIANT 248 248 Q -> R.  
 FT NON-TER 393 393  
 SO SEQUENCE 393 AA: 43684 MW; 239818A9 CRC32;  
 Query Match 100.0%; Score 58; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1.82e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 361 GSRHSSHL 369  
 OY 1 GSRHSSHL 9  
 RESULT 9  
 ID 016809 PRELIMINARY; PRT; 393 AA.  
 AC 016809;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY

CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL: X60019; G506451; -

DR PROSITE: PS00348; P53; 1.

KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.

FT VARIANT 213 213 Q -> R.

FT NON\_TER 393 393

SO SEQUENCE 393 AA: 43694 MW; CB70BD7F CRC32;

Query Match 100.0%; Score 58; DB 2; Length 393;

Best Local Similarity 100.0%; Pred. No. 1.82e-04; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Indels 0; Gaps 0;

Db 361 GSRAHSHL 369

OY 1 GSRAHSHL 9

RESULT 10 PRELIMINARY; PRT; 393 AA.

ID Q15088; PRELIMINARY; PRT; 393 AA.

AC Q15088; PRELIMINARY; PRT; 393 AA.

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

GN P53.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 92007731.

RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;

RL EMBL J. 10:2879-2887(1991).

DR EMBL: X60016; G506445; -

FT VARIANT 238 238 Y -> C.

FT NON\_TER 393 393

SO SEQUENCE 393 AA: 43713 MW; A01E1523 CRC32;

Query Match 100.0%; Score 58; DB 2; Length 393;

Best Local Similarity 100.0%; Pred. No. 1.82e-04; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Indels 0; Gaps 0;

Db 361 GSRAHSHL 369

OY 1 GSRAHSHL 9

RESULT 11 PRELIMINARY; PRT; 393 AA.

ID Q15087; PRELIMINARY; PRT; 393 AA.

AC Q15087; PRELIMINARY; PRT; 393 AA.

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

GN P53.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 92007731.

RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;

RL EMBL J. 10:2879-2887(1991).

DR EMBL: X60014; G506441; -

FT VARIANT 237 237 I -> M.

FT NON\_TER 393 393

SO SEQUENCE 393 AA: 43694 MW; 9BB81992 CRC32;

Query Match 100.0%; Score 58; DB 2; Length 393;

Best Local Similarity 100.0%; Pred. No. 1.82e-04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRAHSHL 369

OY 1 GSRAHSHL 9

RESULT 12 PRELIMINARY; PRT; 301 AA.

ID Q53796; PRELIMINARY; PRT; 301 AA.

AC Q53796; PRELIMINARY; PRT; 301 AA.

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DE BLOWNIN ACETYLTRANSFERASE.

OS STREPTOMYCES VERTICILLIUS.

OC PROKARYOTA; BACTERIA; FIRMICUTES; STREPTOMYCETACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 95129853.

RA CALCUTT M.J., SCHMIDT F.J.;

RL GENE 151:17-21(1994).

DR EMBL: L26955; G434714; -

FT TRANSFERASE.

SO SEQUENCE 301 AA: 32225 MW; C8A31F16 CRC32;

Query Match 75.9%; Score 44; DB 9; Length 301;

Best Local Similarity 62.5%; Pred. No. 1.73e+00;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 5 PRATHL 12

OY 2 PRATHL 9

RESULT 13 PRELIMINARY; PRT; 201 AA.

ID Q16571; PRELIMINARY; PRT; 201 AA.

AC Q16571; PRELIMINARY; PRT; 201 AA.

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)

DE ALPHA-1-ACID GLYCOPROTEIN.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 88029318.

RA DENTE L., PIRAZA M.G., METSPALD A., CORTESE R.;

RL EMBL J. 6:2289-2296(1987).

DR EMBL: M20615; G177861; -

DR EMBL: X06674; G177861; JOINED.

DR EMBL: M20611; G177861; JOINED.

DR EMBL: M20612; G177861; JOINED.

DR EMBL: M20613; G177861; JOINED.

DR EMBL: M20614; G177861; JOINED.

DR EMBL: M20610; G459803; -

DR EMBL: X06675; G459803; JOINED.

DR EMBL: M20606; G459803; JOINED.

DR EMBL: M20607; G459803; JOINED.

DR EMBL: M20608; G459803; JOINED.

DR EMBL: M20609; G459803; JOINED.

SO SEQUENCE 201 AA: 23457 MW; 51896A73 CRC32;

Query Match 74.1%; Score 43; DB 2; Length 201;

Best Local Similarity 55.6%; Pred. No. 3.14e+00;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 141 GGRHVAHL 119

OY 1 GSRAHSHL 9

```

RESULT 14
ID 029495 PRELIMINARY: PRT: 207 AA.
AC 029495:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE ARALKYLAMINE N-ACETYLTRANSFERASE (EC 2.3.1.87)
DE (ARALKYLAMINE N-ACETYLTRANSFERASE) (SEROTONIN ACETYLTRANSFERASE)
DE (SEROTONIN ACETYLASE).
OS OVIS ARLES (SHEEP).
OC: EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC: EUTHERIA; ARTIODACTYLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DORSETT X RAMBOUILLET;
RX MEDLINE: 96099405.
RA COON S.L., ROSEBOOM P.H., BALER R., WELLER J.L., NAMBOODIRI M.A.A.,
RA KOONIN E.V., KLEIN D.C.;
RL SCIENCE 270:1681-1683(1995).
CC -I- CATALYTIC ACTIVITY: ACETYL-COA + ARALKYLAMINE =
CC COA + N-ACETYLARALKYLAMINE.
DR EMBL: U29663; G112944; -
KW TRANSFERASE; ACYLTRANSFERASE.
SQ SEQUENCE 207 AA; 23076 MW; 2B4429A1 CRC32;

```

```

Query Match 74.18; Score 43; DB 4; Length 207;
Best Local Similarity 62.58; Pred. No. 3.14e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

DB 114 PRGSAHL 121
OY 2 SRAHSHL 9

```

```

RESULT 15
ID P89003 PRELIMINARY: PRT: 286 AA.
AC P89003:
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE P53 (FRAGMENT).
OS MASTOMYS NATALENSIS PAPILLOMAVIRUS (MNPV).
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RA LUQUE E.A., TANG L.H., MODLIN I.M.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U48617; G1813453; -
FT NON_TER 1
SQ SEQUENCE 286 AA; 32287 MW; 30F7C9FA CRC32;

```

```

Query Match 72.4%; Score 42; DB 11; Length 286;
Best Local Similarity 87.58; Pred. No. 5.65e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

DB 255 SRAHSHL 262
OY 2 SRAHSHL 9

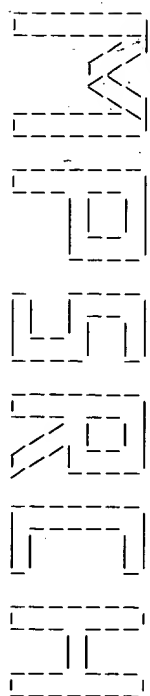
```

```

Search completed: Fri Sep 11 13:57:51 1998
Job time: 23 secs.

```

**THIS PAGE BLANK (USPTO)**



(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*  
MPSrch\_DP protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:51:41 1998; Maspar time 3.42 Seconds

Tabular output not generated. 47.253 Million cell updates/sec

Title: >US-08-452-843-20  
Description: (1-10) from US08452843.pep  
Perfect Score: 78  
Sequence: 1 KPIDGEYFTL 10

Scoring table:  
PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

a.genseq32  
1:part1 2:part2 3:part3 4:part4 5:parts 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 17.694; Variance 50.306; scale 0.352

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description           | Pred. No. |
|------------|-------|-------------|--------|----|--------|-----------------------|-----------|
| 1          | 78    | 100.0       | 10     | 18 | R89381 | p53 derived immunogen | 3.62e-02  |
| 2          | 78    | 100.0       | 48     | 23 | W20035 | Human p53 tetramerisa | 3.62e-02  |
| 3          | 78    | 100.0       | 73     | 26 | W22024 | Wild-type p53 tetrame | 3.62e-02  |
| 4          | 78    | 100.0       | 74     | 21 | W09322 | C-terminal domain of  | 3.62e-02  |
| 5          | 78    | 100.0       | 113    | 10 | R51877 | Human p53 amino acids | 3.62e-02  |
| 6          | 78    | 100.0       | 157    | 10 | R51878 | Human p53 amino acids | 3.62e-02  |
| 7          | 78    | 100.0       | 310    | 10 | R51873 | Human p53 amino acids | 3.62e-02  |
| 8          | 78    | 100.0       | 328    | 10 | R51876 | Human p53 amino acids | 3.62e-02  |
| 9          | 78    | 100.0       | 354    | 10 | R51874 | Human p53 amino acids | 3.62e-02  |
| 10         | 78    | 100.0       | 363    | 21 | W13976 | Modified p53 variant  | 3.62e-02  |
| 11         | 78    | 100.0       | 363    | 21 | W13971 | Modified p53 variant  | 3.62e-02  |
| 12         | 78    | 100.0       | 363    | 21 | W13974 | Modified p53 variant  | 3.62e-02  |
| 13         | 78    | 100.0       | 363    | 21 | W13973 | Modified p53 variant  | 3.62e-02  |
| 14         | 78    | 100.0       | 363    | 21 | W13972 | Modified p53 variant  | 3.62e-02  |
| 15         | 78    | 100.0       | 363    | 21 | W13975 | Modified p53 variant  | 3.62e-02  |
| 16         | 78    | 100.0       | 363    | 21 | W13977 | Modified p53 variant  | 3.62e-02  |
| 17         | 78    | 100.0       | 368    | 21 | W13956 | Chimeric p53 protein. | 3.62e-02  |
| 18         | 78    | 100.0       | 370    | 21 | W13957 | Chimeric p53 protein. | 3.62e-02  |

|    |    |       |     |    |        |                       |          |
|----|----|-------|-----|----|--------|-----------------------|----------|
| 19 | 78 | 100.0 | 374 | 24 | W28482 | Human p53 protein var | 3.62e-02 |
| 20 | 78 | 100.0 | 374 | 24 | W28481 | Human p53 protein var | 3.62e-02 |
| 21 | 78 | 100.0 | 381 | 24 | W28490 | Human p53 protein var | 3.62e-02 |
| 22 | 78 | 100.0 | 381 | 24 | W28489 | Human p53 protein var | 3.62e-02 |
| 23 | 78 | 100.0 | 390 | 19 | W02623 | Mouse p53 protein.    | 3.62e-02 |
| 24 | 78 | 100.0 | 393 | 19 | W02617 | Human p53 tumour supp | 3.62e-02 |
| 25 | 78 | 100.0 | 393 | 21 | W13970 | Modified p53 variant  | 3.62e-02 |
| 26 | 78 | 100.0 | 393 | 21 | W13968 | Modified p53 variant  | 3.62e-02 |
| 27 | 78 | 100.0 | 393 | 21 | W25155 | Human p53 variant fou | 3.62e-02 |
| 28 | 78 | 100.0 | 393 | 22 | W13953 | Human p53 variant fou | 3.62e-02 |
| 29 | 78 | 100.0 | 393 | 22 | W13952 | Human p53 variant fou | 3.62e-02 |
| 30 | 78 | 100.0 | 393 | 22 | W13979 | Human tumour-derived  | 3.62e-02 |
| 31 | 78 | 100.0 | 393 | 21 | W05345 | Human tumour-derived  | 3.62e-02 |
| 32 | 78 | 100.0 | 393 | 22 | W13951 | Human p53 mutant N239 | 3.62e-02 |
| 33 | 78 | 100.0 | 393 | 14 | R79658 | Human tumour-derived  | 3.62e-02 |
| 34 | 78 | 100.0 | 393 | 22 | W13948 | Human p53 protein.    | 3.62e-02 |
| 35 | 78 | 100.0 | 393 | 16 | R94623 | Human wild-type p53 t | 3.62e-02 |
| 36 | 78 | 100.0 | 393 | 5  | R22238 | p53 protein.          | 3.62e-02 |
| 37 | 78 | 100.0 | 393 | 5  | R26758 | Sequence of 53 kd cel | 3.62e-02 |
| 38 | 78 | 100.0 | 393 | 18 | R91933 | p53.                  | 3.62e-02 |
| 39 | 78 | 100.0 | 393 | 21 | W05346 | Wild type p53 protein | 3.62e-02 |
| 40 | 78 | 100.0 | 401 | 24 | W28487 | Human p53 mutant R273 | 3.62e-02 |
| 41 | 78 | 100.0 | 401 | 24 | W28488 | Human p53 protein var | 3.62e-02 |
| 42 | 78 | 100.0 | 411 | 21 | W13967 | Human p53 protein var | 3.62e-02 |
| 43 | 78 | 100.0 | 438 | 14 | R74272 | Chimeric p53 protein. | 3.62e-02 |
| 44 | 78 | 100.0 | 438 | 10 | R50088 | Tumour suppressor pro | 3.62e-02 |
| 45 | 78 | 100.0 | 533 | 23 | W19763 | p53 tumour suppressor | 3.62e-02 |
|    |    |       |     |    |        | p53-GM-CSF immunostim | 3.62e-02 |

## ALIGNMENTS

RESULT 1  
ID R89381 standard; peptide: 10 AA.  
AC R89381;  
DE 18-SEP-1996 (first entry)  
DE p53 derived immunogenic peptide, residues 321-330.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN W09603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; W09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI; 96-116784/12.  
PT Compsn. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PT Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g. the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 10 AA:

Query Match 100.0%; Score 78; DB 18; Length 10;  
Best Local Similarity 100.0%; Pred. No. 3.62e-02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 1 KPIDGEYFTL 10  
QY 1 KPIDGEYFTL 10

RESULT 2  
ID W20035 standard: Protein: 48 AA.  
AC W20035:  
DE 04-SEP-1997 (first entry)  
KW Human p53 tetramerisation domain flanked by linking sequences.  
KW Multimerisation: self assembly; functional linker; folding;  
KW multimerisation domain; post-translational modification; secretion;  
KW Interleukin-2; TAF1131; TAF180; TATA box binding associated factor;  
KW p53; histone H3; H4; thrombospondin; TSP-4; platelet factor; PF4;  
KW cartilage oligomeric protein; COMP.  
OS Synthetic.  
FT Key location/qualifiers  
FT domain 3.43  
FT /note="human p53 tetramerisation domain"  
PN W09637621-A2.  
PD 28-NOV-1996.  
PE 23-MAY-1996; E02230.  
PR 23-MAY-1995; EP-107914.  
PA (MORP-) MORPHOSYS GRS PROTEINOPTIMIERUNG MBH.  
PI Hoess A, Pack F.  
FI WPI: 97-021226/02.  
DR N-PSDB: T71287.  
PT Multimerisation devices for self assembly of multifunctional  
PT proteins - used to express recombinant multivalent polypeptide(s)  
PT by incorporation in a cistron encoding the protein  
PS Claim 3; Fig 3; 64pp; English.  
CC W20035 is the product of an expression cassette encoding the  
CC tetramerisation domain of human p53 (residues 319-360). The cassette was  
CC incorporated into a larger DNA sequence comprising, 5' to 3', a 1st  
CC functional domain: a 1st linker sequence; a multimerisation device;  
CC a 2nd linker sequence; and a 2nd functional domain. The multimerisation  
CC device allows the combination of two or more functional domains in a  
CC structure which is capable of self-multimerisation (at least  
CC trimerisation). Functional domains may, for example, bind to a defined  
CC target, catalyse a reaction, block a receptor binding site, inhibit the  
CC action of another protein or bind to a metal ion. Multimerisation  
CC domains from p53, platelet factor 4, thrombospondin, TSP-4, TATA box  
CC binding associated factors and cartilage oligomeric protein may be  
CC used. The multifunctional proteins can be prepared using standard  
CC recombinant micro-organisms, even though the molecular weight of the  
CC assembled protein exceeds that of the proteins commonly expressed in  
CC bacteria. They have low immunogenicity in humans and carry two or more  
CC functions in a single multimeric structure. Use of a combination of in  
CC vivo expression and in vitro synthesis overcomes prior art problems  
CC due to the differences in folding, secretion and post-translational  
CC modifications for different polypeptides in different hosts.  
SQ Sequence 48 AA:  
Query Match 100.0%; Score 78; DB 23; Length 48;  
Best Local Similarity 100.0%; Pred. No. 3.62e-02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PI Anthony-Cahill SJ, Epp JK, Kevin BA, Mathews AJ;  
FI Olin PO;  
DR WPI: 97-351067/32.  
PT New globin containing non-natural binding site and related nucleic  
PT acid - also multimeric haemoglobin, used as oxygen carrier for in  
PT vivo or in vitro applications, with extended half-life and reduced  
PT extravasation  
PS Example 8; Page 40; 64pp; English.  
CC This sequence represents the tetramerising domain of p53. This sequence,  
CC or the oligomerising domains of the yeast transcription factor GCN4 (see  
CC W22019 and W22020) can be used in the globin of the invention. The  
CC globin of the invention has a non-natural binding domain (BD), preferably  
CC an oligomerising domain or ligand binding domain. The globin may be  
CC combined with other globins to form a multimeric haemoglobin (Hb). The Hb  
CC are used as oxygen carriers, both in vivo (as blood substitutes, volume  
CC extenders, in treatment of anaemia and to stimulate haematopoiesis) and  
CC in vitro (e.g. to improve growth of cell cultures). They are also used to  
CC remove oxygen from solutions, or therapeutically to remove nitric oxide.  
CC The Hb can also be used as a reference standard for analytical  
CC instruments and for delivering drugs or in in vivo imaging. Incorporation  
CC of the BD allows production of larger Hb that can be assembled without  
CC using exogenous crosslinking agents, and the size of the multimer can be  
CC controlled. Large Hb show reduced extravasation and prolonged half life,  
CC and are able to deliver oxygen to tissues which erythrocytes can not  
CC reach (e.g. downstream of a thrombus, angioplasty balloon etc.).  
SQ Sequence 73 AA:  
Query Match 100.0%; Score 78; DB 26; Length 73;  
Best Local Similarity 100.0%; Pred. No. 3.62e-02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 27 kpldgyftl 36  
QY 1 KPLDGYFTL 10

RESULT 4  
ID W09322 standard: peptide: 74 AA.  
AC W09322:  
DE 10-JUN-1997 (first entry)  
KW C-terminal domain of p53 protein.  
KW Chimeric: bispecific; DNA binding domain; trans: activator; repressor;  
KW diphtheria; pseudomonas; toxin; thymidine kinase; single chain antibody;  
KW pathogen; HIV tat; papilloma virus; E6/E7; Epstein-Barr virus; EBNA;  
KW hyperproliferation; p53; tumour; oligomerisation.  
OS Homo sapiens.  
PN W09630512-A1.  
PD 03-OCT-1996.  
PE 29-MAR-1996; F00477.  
PR 31-MAR-1995; FR-003841.  
PA (RHON) RHONE POULENC RORER SA.  
PI Bracco L, Schweighoffer F, Tocque B;  
DR WPI: 96-45359/45.  
PT Conditional gene expression system triggered by e.g. infection or  
PT hyper-proliferation - comprises novel bispecific proteins having  
PT DNA-binding domain and second domain specific for trans-activator or  
PT repressor, for gene therapy.  
PS Claim 16; Page 44; 81pp; French.  
CC The invention relates to novel chimeric, bispecific proteins which  
CC comprise: (a) a DNA binding domain and (b) a domain which binds a  
CC trans-activator (TA), trans-repressor (TR) or their complexes, which are  
CC characteristic of a physiological or pathophysiological state. The novel  
CC chimeric, bispecific proteins allow expression of a therapeutic protein  
CC (e.g. diphtheria or Pseudomonas toxins, thymidine kinase, single chain  
CC antibodies) to be regulated in response to particular conditions.  
CC Examples include making the protein responsive to the presence of  
CC particular pathogenic TA mols (e.g. HIV tat, papilloma virus E6/E7  
CC proteins or Epstein-Barr virus EBNA protein), the therapeutic protein  
CC will be expressed in those cells infected by that pathogen. Similarly,  
CC where the chimeric protein responds to a cellular protein typical of a  
CC hyperproliferative state (esp. wild-type and mutant p53), expression can  
CC be restricted to tumour cells. The sequence presented here is an example  
CC of a TA binding domain. It corresponds to the C-terminal domain of the

CC p53 protein between residues 320-393 containing the oligomerisation  
 CC domain which binds TA proteins.  
 SO Sequence 74 AA;

Query Match 100.0%; Score 78; DB 21; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 3.62e-02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 2 kpldgyefl 11  
 |||||  
 1 kpldgyefl 10

RESULT 5  
 ID R51877 standard; Protein: 113 AA.  
 AC R51877;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 237-349.  
 KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 KW anti-oncogene; cancer; tumour; antibody binding region; epitope.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 37 /note="Arg corresponds to a CAT codon"

PN WO9408241-A.  
 PD 14-APR-1994.  
 PF 30-SEP-1993; E02666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schranz P, Tessmer C, Volkmann M, Zentgraf H;  
 DR WPI: 94-135732/16.  
 DR N-PSDB: 062362.  
 PT Non-radioactive detection of p53 specific antibodies - by capture  
 PT on immobilised p53 or its fragments, then reaction with labelled  
 PT second antibody, for diagnosis of tumours and suitable for  
 PT screening  
 PS Claim 10: Page 19; 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 SO Sequence 113 AA;

Query Match 100.0%; Score 78; DB 10; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 3.62e-02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 85 kpldgyefl 94  
 |||||  
 1 kpldgyefl 10

RESULT 6  
 ID R51878 standard; Protein: 157 AA.  
 AC R51878;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 237-393.  
 KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 KW anti-oncogene; cancer; tumour; antibody binding region; epitope.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 37 /note="Arg corresponds to a CAT codon"  
 PN WO9408241-A.  
 PD 14-APR-1994.  
 PF 30-SEP-1993; E02666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schranz P, Tessmer C, Volkmann M, Zentgraf H;  
 DR WPI: 94-135732/16.  
 DR N-PSDB: 062363.  
 PT Non-radioactive detection of p53 specific antibodies - by capture

PT on immobilised p53 or its fragments, then reaction with labelled  
 PT second antibody, for diagnosis of tumours and suitable for  
 PT screening  
 PS Claim 10: Page 19; 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 SO Sequence 157 AA;

Query Match 100.0%; Score 78; DB 10; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 3.62e-02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 85 kpldgyefl 94  
 |||||  
 1 kpldgyefl 10

RESULT 7  
 ID R51873 standard; Protein: 310 AA.  
 AC R51873;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 40-349.  
 KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 KW anti-oncogene; cancer; tumour; antibody binding region; epitope.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 234 /note="Arg corresponds to a CAT codon"  
 PN WO9408241-A.  
 PD 14-APR-1994.  
 PF 30-SEP-1993; E02666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schranz P, Tessmer C, Volkmann M, Zentgraf H;  
 DR WPI: 94-135732/16.  
 DR N-PSDB: 062358.  
 PT Non-radioactive detection of p53 specific antibodies - by capture  
 PT on immobilised p53 or its fragments, then reaction with labelled  
 PT second antibody, for diagnosis of tumours and suitable for  
 PT screening  
 PS Claim 10: Page 17; 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 SO Sequence 310 AA;

Query Match 100.0%; Score 78; DB 10; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 3.62e-02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 282 kpldgyefl 291  
 |||||  
 1 kpldgyefl 10

RESULT 8  
 ID R51876 standard; Protein: 328 AA.  
 AC R51876;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 66-393.  
 KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 KW anti-oncogene; cancer; tumour; antibody binding region; epitope.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 208 /note="Arg corresponds to a CAT codon"  
 PN WO9408241-A.  
 PD 14-APR-1994.

PE 30-SEP-1993; E02666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schranz P, Tesserer C, Volkmann M, Zentgraf H;  
 DR WPI: 94-135732/16.  
 DR N-PSDB: 062361.  
 PT Non-radioactive detection of p53 specific antibodies - by capture  
 on immobilised p53 or its fragments, then reaction with labelled  
 PT second antibody, for diagnosis of tumours and suitable for  
 screening  
 PS Claim 10; Page 18; 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 SQ Sequence 328 AA;  
 Query Match 100.0%; Score 78; DB 10; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 3,62e-02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 256 kpldgyftl 265  
 OY 1 KPLDGEYFTL 10  
 RESULT 9  
 ID R51874 standard; Protein: 354 AA.  
 AC R51874;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 40-393.  
 KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 OS anti-oncogene; cancer; tumour; antibody binding region; epitope.  
 FH Homo sapiens.  
 FT Key / Location/Qualifiers  
 FT misc\_difference 234  
 FT /note="Arg corresponds to a CAT codon"  
 PN MO9408241-A.  
 PD 14-APR-1994.  
 PF 30-SEP-1993; E02666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schranz P, Tesserer C, Volkmann M, Zentgraf H;  
 DR WPI: 94-135732/16.  
 DR N-PSDB: 062359.  
 PT Non-radioactive detection of p53 specific antibodies - by capture  
 on immobilised p53 or its fragments, then reaction with labelled  
 PT second antibody, for diagnosis of tumours and suitable for  
 screening  
 PS Claim 10; Page 18; 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 SQ Sequence 354 AA;  
 Query Match 100.0%; Score 78; DB 10; Length 354;  
 Best Local Similarity 100.0%; Pred. No. 3,62e-02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 282 kpldgyftl 291  
 OY 1 KPLDGEYFTL 10  
 RESULT 10  
 ID W13976 standard; Protein: 363 AA.  
 AC W13976;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53C273del1364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 PR 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53C273del1364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;

KW Apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN W09710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INSR ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1; 59-61; 82pp; English.  
 CC Modified p53 variant p53C273del1364-393 (W13976) has the tumour-  
 CC derived cysteine 273 mutation (see also W13952) and a deletion  
 CC of the C-terminal 30 amino acids of wild-type p53 (see also  
 CC W13948). Cys273 is a Class I p53 tumour mutation that affects DNA  
 CC binding. The C-terminal deletion, introduced by site-directed  
 CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
 CC tumour mutant. This provides the means for pharmacological rescue  
 CC of p53 function in cancer patients. Other modified p53 constructs  
 CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
 CC acids coding for modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;  
 Query Match 100.0%; Score 78; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 3,62e-02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 kpldgyftl 330  
 OY 1 KPLDGEYFTL 10  
 RESULT 11  
 ID W13971 standard; Protein: 363 AA.  
 AC W13971;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53R284del1364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW Apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN W09710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INSR ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1; 51-52; 82pp; English.  
 CC Modified p53 variant p53R284del1364-393 (W13971) has a Thr284 to Arg  
 CC substn. (see also W13949) and a deletion of the C-terminal 30  
 CC amino acids. The R284R substitution, introduced by site-directed  
 CC mutagenesis of p53 DNA, provides a novel p53-DNA contact between a  
 CC phosphate of the DNA backbone and p53. The C-terminal deletion  
 CC permits in vitro DNA binding. The variant provides the means for  
 CC pharmacological rescue of p53 function in cancer patients. Other  
 CC modified p53 constructs (W13949-50, W13953-54, W13968-77) have also  
 CC been produced. Nucleic acids coding for modified p53 can be used  
 CC for cancer gene therapy.  
 SQ Sequence 363 AA;  
 Query Match 100.0%; Score 78; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 3,62e-02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 kpldgyftl 330  
 OY 1 KPLDGEYFTL 10



RESULT 12

ID W13974 standard; Protein; 363 AA.

AC W13974:

DE 25-JUN-1997 (first entry)

DE Modified p53 variant p53H273R284del1364-393.

KW p53; tumour suppressor; cancer; therapy; cell proliferation;

KW apoptosis; protein engineering; DNA binding.

OS Synthetic.

PN W09710843-A1.

PD 27-MAR-1997.

PF 20-SEP-1996; U15188.

PR 22-SEP-1995; US-004802.

PR 21-AUG-1996; US-697221.

PA (W1ST-) W1STAR INST ANATOMY & BIOLOGY.

PI Halazonetis TD.

DR WPI: 97-202618/18.

PT R284K modified p53 protein having DNA binding ability - useful in treatment of cancer

PS Example 1: 56-57; 82pp; English.

CC Modified p53 variant p53H273R284del1364-393 (W13974) has the tumour-derived histidine 273 mutation (see also W13952) and a deletion of the C-terminal 30 amino acids of wild-type p53 (see also W13948). His273 is a Class I p53 tumour mutation that affects DNA binding. The C-terminal deletion, introduced by site-directed mutagenesis of p53 DNA, activates the DNA binding of the p53 tumour mutant. This provides the means for pharmacological rescue of p53 function in cancer patients. Other modified p53 constructs (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic acids coding for modified p53 can be used for cancer gene therapy.

SQ Sequence 363 AA;

Query Match 100.0%; Score 78; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 3.62e-02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 kpldgyftl 330

QY 1 kpldgyftl 10

RESULT 13

ID W13973 standard; Protein; 363 AA.

AC W13973:

DE 25-JUN-1997 (first entry)

DE Modified p53 variant p53Q248R284del1364-393.

KW p53; tumour suppressor; cancer; therapy; cell proliferation;

KW apoptosis; protein engineering; DNA binding.

OS Synthetic.

PN W09710843-A1.

PD 27-MAR-1997.

PF 20-SEP-1996; U15188.

PR 22-SEP-1995; US-004802.

PR 21-AUG-1996; US-697221.

PA (W1ST-) W1STAR INST ANATOMY & BIOLOGY.

PI Halazonetis TD.

DR WPI: 97-202618/18.

PT R284K modified p53 protein having DNA binding ability - useful in treatment of cancer

PS Example 1: 54-56; 82pp; English.

CC Modified p53 variant p53Q248R284del1364-393 (W13973) has the tumour-derived Glu248 mutation (see also W13951), a Thr284 to Arg substitution (see also W13949) and a deletion of the 30 C-terminal amino acids of wild-type p53 (W13948). Glu248 is a Class I p53 tumour mutation that affects DNA binding. The R284R substitution, introduced by site-directed mutagenesis of p53 DNA, provides a novel p53-DNA contact between a phosphate of the DNA backbone and p53, and restores DNA binding. The C-terminal deletion permits in vitro DNA binding. The construct provides the means for pharmacological rescue of p53 function in cancer patients. Other modified p53 constructs (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic acids coding for modified p53 can be used for cancer gene therapy.

SQ Sequence 363 AA;

Query Match 100.0%; Score 78; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 3.62e-02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 kpldgyftl 330

QY 1 kpldgyftl 10

RESULT 14

ID W13972 standard; Protein; 363 AA.

AC W13972:

DE 25-JUN-1997 (first entry)

DE Modified p53 variant p53Q248del1364-393.

KW p53; tumour suppressor; cancer; therapy; cell proliferation;

KW apoptosis; protein engineering; DNA binding.

OS Synthetic.

PN W09710843-A1.

PD 27-MAR-1997.

PF 20-SEP-1996; U15188.

PR 22-SEP-1995; US-004802.

PR 21-AUG-1996; US-697221.

PA (W1ST-) W1STAR INST ANATOMY & BIOLOGY.

PI Halazonetis TD.

DR WPI: 97-202618/18.

PT R284K modified p53 protein having DNA binding ability - useful in treatment of cancer

PS Example 1: 53-54; 82pp; English.

CC Modified p53 variant p53Q248del1364-393 (W13972) has the tumour-derived glutamine 248 mutation (see also W13951) and a deletion of the C-terminal 30 amino acids of wild-type p53 (see also W13948). Glu248 is a Class I p53 tumour mutation that affects DNA binding. The C-terminal deletion, introduced by site-directed mutagenesis of p53 DNA, activates the DNA binding of the p53 tumour mutant. This provides the means for pharmacological rescue of p53 function in cancer patients. Other modified p53 constructs (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic acids coding for modified p53 can be used for cancer gene therapy.

SQ Sequence 363 AA;

Query Match 100.0%; Score 78; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 3.62e-02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 kpldgyftl 330

QY 1 kpldgyftl 10

RESULT 15

ID W13975 standard; Protein; 363 AA.

AC W13975:

DE 25-JUN-1997 (first entry)

DE Modified p53 variant p53H273R284del1364-393.

KW p53; tumour suppressor; cancer; therapy; cell proliferation;

KW apoptosis; protein engineering; DNA binding.

OS Synthetic.

PN W09710843-A1.

PD 27-MAR-1997.

PF 20-SEP-1996; U15188.

PR 22-SEP-1995; US-004802.

PR 21-AUG-1996; US-697221.

PA (W1ST-) W1STAR INST ANATOMY & BIOLOGY.

PI Halazonetis TD.

DR WPI: 97-202618/18.

PT R284K modified p53 protein having DNA binding ability - useful in treatment of cancer

PS Example 1: 58-59; 82pp; English.

CC Modified p53 variant p53H273R284del1364-393 (W13975) has the tumour-derived His273 mutation (see also W13952), a Thr284 to Arg substitution (see also W13949) and a deletion of the 30 C-terminal amino acids

CC of wild-type p53 (W13948). His273 is a Class I p53 tumour mutation  
CC that affects DNA binding. The T284R substitution, introduced by  
CC site-directed mutagenesis of p53 DNA, provides a novel p53-DNA  
CC contact between a phosphate of the DNA backbone and p53, and  
CC restores DNA binding. The C-terminal deletion permits in vitro  
CC DNA binding. The construct provides the means for pharmacological  
CC rescue of p53 function in cancer patients. Other modified p53  
CC constructs (W13949-50, W13953-54, W13968-77) have also been  
CC produced. Nucleic acids coding for modified p53 can be used for  
CC cancer gene therapy.

SO Sequence 363 AA;

Query Match 100.0%; Score 78; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 3.62e-02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 kpldageyftl 330

QY 1 kpldageyftl 10

Search completed: Fri Sep 11 13:51:57 1998  
Job time : 16 secs.

\*\*\*\*\*  
 WIRETH  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*  
 Msrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Sep 11 13:52:14 1998; MasPar time 3.31 Seconds  
 Tabular output not generated. 110.293 Million cell updates/sec

Title: >US-08-452-843-20  
 Description: (1-10) from US08452843.pep  
 Perfect Score: 78  
 Sequence: 1 KPIDGEXFTL 10

Scoring table:  
 PAM 150  
 Gap 15

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: p1r56  
 1:p1r1 2:p1r2 3:p1r3 4:p1r4 5:nr13d

Statistics: Mean 24.771; Variance 34.958; scale 0.709

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No.            |
|------------|-------|-------------|--------|----|-------------|----------------------|
| 1          | 78    | 100.0       | 42     | 5  | 1SAIB       | tumor suppressor p53 |
| 2          | 78    | 100.0       | 42     | 5  | 1SAIA       | tumor suppressor p53 |
| 3          | 78    | 100.0       | 42     | 5  | 1SAID       | tumor suppressor p53 |
| 4          | 78    | 100.0       | 42     | 5  | 1SAIC       | tumor suppressor p53 |
| 5          | 78    | 100.0       | 42     | 5  | 1SAHA       | tumor suppressor p53 |
| 6          | 78    | 100.0       | 42     | 5  | 1SAHB       | tumor suppressor p53 |
| 7          | 78    | 100.0       | 42     | 5  | 1SAHC       | tumor suppressor p53 |
| 8          | 78    | 100.0       | 42     | 5  | 1SAHD       | tumor suppressor p53 |
| 9          | 78    | 100.0       | 42     | 5  | 1SAGD       | tumor suppressor p53 |
| 10         | 78    | 100.0       | 42     | 5  | 1SAGC       | tumor suppressor p53 |
| 11         | 78    | 100.0       | 42     | 5  | 1SAGB       | tumor suppressor p53 |
| 12         | 78    | 100.0       | 42     | 5  | 1SAGA       | tumor suppressor p53 |
| 13         | 78    | 100.0       | 42     | 5  | 1SAFC       | tumor suppressor p53 |
| 14         | 78    | 100.0       | 42     | 5  | 1SAFD       | tumor suppressor p53 |
| 15         | 78    | 100.0       | 42     | 5  | 1SAFA       | tumor suppressor p53 |
| 16         | 78    | 100.0       | 42     | 5  | 1SAFB       | tumor suppressor p53 |
| 17         | 78    | 100.0       | 42     | 5  | 1SAJB       | tumor suppressor p53 |
| 18         | 78    | 100.0       | 42     | 5  | 1SAJA       | tumor suppressor p53 |
| 19         | 78    | 100.0       | 42     | 5  | 1SAJC       | tumor suppressor p53 |
| 20         | 78    | 100.0       | 42     | 5  | 1SAJD       | tumor suppressor p53 |
| 21         | 78    | 100.0       | 42     | 5  | 1SAJE       | tumor suppressor p53 |
| 22         | 78    | 100.0       | 42     | 5  | 1SAEB       | tumor suppressor p53 |
| 23         | 78    | 100.0       | 42     | 5  | 1SAEC       | tumor suppressor p53 |

| RESULT    | ENTRY   | 1              | 24 | 25    | 26  | 27 | 28     | 29                     | 30       | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 |
|-----------|---|----------------|----|-------|-----|----|--------|------------------------|----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| TITLE     | 1SAIB   | #type complete | 78 | 100.0 | 42  | 5  | 1SAED  | tumor suppressor p53   | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| PDB_TITLE | tumor suppressor p53 oligomerization domain, residues 319-360   |                | 78 | 100.0 | 42  | 5  | 1SALD  | tumor suppressor p53   | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| ORGANISM  | human   |                | 78 | 100.0 | 42  | 5  | 1SALC  | tumor suppressor p53   | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | high resolution solution NMR structure of the oligomerization domain of p53 by multi-dimensional NMR (sac structures) |                | 78 | 100.0 | 42  | 5  | 1SALA  | tumor suppressor p53   | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 42  | 5  | 1SALB  | tumor suppressor p53   | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #cross-references   |                | 78 | 100.0 | 42  | 5  | 1OLGC  | p53 (oligomerization)  | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 42  | 5  | 1OLGB  | p53 (oligomerization)  | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 42  | 5  | 1SAEC  | tumor suppressor p53   | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 42  | 5  | 1OLGA  | p53 (oligomerization)  | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 42  | 5  | 1OLGD  | p53 (oligomerization)  | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 42  | 5  | 1OLGC  | p53 (oligomerization)  | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 42  | 5  | 1OLGB  | p53 (oligomerization)  | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 42  | 5  | 1SAKB  | tumor suppressor p53   | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 42  | 5  | 1SAKD  | tumor suppressor p53   | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 386 | 2  | SS1648 | cellular tumor antigen | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 390 | 1  | DNMS53 | cellular tumor antigen | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 391 | 2  | SO2192 | cellular tumor antigen | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 393 | 1  | DNRH53 | cellular tumor antigen | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| REFERENCE | #authors  |                | 78 | 100.0 | 393 | 2  | SO6594 | cellular tumor antigen | 1.18e-04 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |

## ALIGNMENTS

RESULT 1  
 ENTRY 1  
 TITLE 1SAIB  
 #type complete  
 tumor suppressor p53 oligomerization domain, residues 319-360  
 sac structures 53-77, chain B - human  
 high resolution solution NMR structure of the oligomerization domain of p53 by multi-dimensional NMR (sac structures)  
 #format\_name Homo sapiens #common\_name man  
 A65596  
 ORGANISM  
 #authors Clore, G.M.; Omichinski, J.G.; Gronenborn, A.M.  
 #submission submitted to the Brookhaven Protein Data Bank, March 1995  
 #cross-references PDB:1SAI  
 TN031001  
 #authors  
 Clore, G.M.; Ernst, J.; Clubb, R.; Omichinski, J.G.;  
 Polnitzer Kennedy, W.M.; Sakaguchi, K.; Appella, E.;  
 Gronenborn, A.M.  
 Nat. Struct. Biol. (1995) 2:321  
 Refined solution structure of the oligomerization domain of  
 the tumour suppressor p53.  
 TN031002  
 #authors  
 Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;  
 Sakamoto, H.; Appella, E.; Gronenborn, A.M.  
 Science (1995) 267:1515  
 Interhelical angles in the solution structure of the  
 oligomerization domain of p53: correction.  
 TN031003  
 #authors  
 Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;  
 Sakamoto, H.; Appella, E.; Gronenborn, A.M.  
 Science (1994) 265:386  
 High-resolution structure of the oligomerization domain of  
 p53 by multidimensional NMR.  
 COMMENT Resolution: not applicable  
 DETERMINATION: NMR  
 FEATURE  
 17-37 #region helix (right hand alpha)\  
 9-15 #region beta sheet  
 SUMMARY #length 42 #molecular\_weight 4940 #checksum 6203  
 Query Match 100.0%; Score 78; DB 5; Length 42;  
 Best Local Similarity 100.0%; Pred. No. 1.18e-04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|                       |  |
|-----------------------|--|
| RESULT                | 2  |
| ENTRY                 |  |
| TITLE                 | ISAIA #type complete   |
| PDB_TITLE             | tumor suppressor p53 oligomerization domain, residues 319-360  |
| ORGANISM              | sac structures 53-77, chain A - human  |
| REFERENCE             | high resolution solution NMR structure of the oligomerization domain of p53 by multi-dimensional NMR (sac structures)            |
| #authors              | #formal_name Homo sapiens #common_name man   |
| #submission           | A66596   |
| #cross-references     | Cloire, G.M.; Omichinski, J.G.; Gronenborn, A.M.   |
| REFERENCE             | submitted to the Brookhaven Protein Data Bank, March 1995  |
| #journal              | #TNO30998  |
| #title                | Cloire, G.M.; Ernst, J.; Ciubb, R.; Omichinski, J.G.;<br>Polinder-Kennedy, W.M.; Sakaguchi, K.; Appella, E.;<br>Gronenborn, A.M. |
| #authors              | Nat. Struct. Biol. (1995) 2:321  |
| REFERENCE             | Refined solution structure of the oligomerization domain of<br>the tumour suppressor p53.<br>TN030999                            |
| #journal              | Cloire, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;<br>Sakamoto, H.; Appella, E.; Gronenborn, A.M.                      |
| #title                | Science (1995) 267:1515  |
| REFERENCE             | Interhelical angles in the solution structure of the<br>oligomerization domain of p53: correction.<br>TN031000                   |
| #authors              | Cloire, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;<br>Sakamoto, H.; Appella, E.; Gronenborn, A.M.                      |
| #journal              | Science (1994) 265:386   |
| #title                | High-resolution structure of the oligomerization domain of<br>p53 by multidimensional NMR.<br>Resolution: not applicable         |
| COMMENT               | Determination: NMR   |
| KEYWORDS:             | anti-oncogene  |
| FEATURE               |  |
| 17-37                 | #region helix (right hand alpha)\  |
| 9-15                  | #region beta sheet   |
| SUMMARY               | #length 42; #molecular-weight 4940 #checksum 6203  |
| Query Match           | 100.0%; Score 78; DB 5; Length 42;   |
| Best Local Similarity | 100.0%; Pred. No. 1,18e+04;  |
| Matches               | 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  |
| DB                    | 3 KPLDGEFTL 12   |
| OY                    | 1 KPLDGEFTL 10   |
| RESULT                | 3  |
| ENTRY                 |  |
| TITLE                 | ISAID #type complete   |
| PDB_TITLE             | tumor suppressor p53 oligomerization domain, residues 319-360  |
| ORGANISM              | sac structures 53-77, chain D - human  |
| REFERENCE             | high resolution solution NMR structure of the oligomerization domain of p53 by multi-dimensional NMR (sac structures)            |
| #authors              | #formal_name Homo sapiens #common_name man   |
| #submission           | A66596   |
| #cross-references     | Cloire, G.M.; Omichinski, J.G.; Gronenborn, A.M.   |
| REFERENCE             | submitted to the Brookhaven Protein Data Bank, March 1995  |
| #journal              | #TNO31007  |
| #title                | Cloire, G.M.; Ernst, J.; Ciubb, R.; Omichinski, J.G.;<br>Polinder-Kennedy, W.M.; Sakaguchi, K.; Appella, E.;<br>Gronenborn, A.M. |
| #authors              | Nat. Struct. Biol. (1995) 2:321  |
| REFERENCE             | Refined solution structure of the oligomerization domain of<br>the tumour suppressor p53.<br>TN031008                            |
| #journal              | Cloire, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;<br>Sakamoto, H.; Appella, E.; Gronenborn, A.M.                      |
| #title                | Science (1995) 267:1515  |
| REFERENCE             | Interhelical angles in the solution structure of the<br>oligomerization domain of p53: correction.<br>TN031009                   |
| #authors              | Cloire, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;   |

```

#journal      Sakamoto, H.; Appella, E.; Gronenborn, A.M.
#title        Science (1994) 265:386
              High-resolution structure of the oligomerization domain of
              p53 by multidimensional NMR.
COMMENT       Resolution: not applicable
COMMENT       Determination: NMR
KEYWORDS      anti-oncogene
FEATURE       17-37
              #region helix (right hand alpha)\
              #region beta sheet
SUMMARY       #length 42 #molecular-weight 4940 #checksum 6203

Query Match   100.0%: Score 78; DB 5; Length 42;
Best Local Similarity 100.0%: Pident. No. 1,18e-04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db            3 KPLDGEYFTL 12
ENTRY         |||||
TITLE         1 KPLDGEYFTL 10

RESULT        4
ENTRY         1SAIC      #type complete
TITLE         tumor suppressor p53 oligomerization domain, residues 319-360
              sac structures 5377, chain C - human
              high resolution solution NMR structure of the oligomerization
              domain of p53 by multi-dimensional NMR (sac structures)
ORGANISM      #formal_name Homo sapiens #common_name man
REFERENCE      A66596
#authors       Clore, G.M.; Omichinski, J.G.; Gronenborn, A.M.
#submission    submitted to the Brookhaven Protein Data Bank, March 1995
#cross-references PDB:1SAI
REFERENCE      TN031004
#authors       Clore, G.M.; Ernst, J.; Clubb, R.; Omichinski, J.G.;
              Poldander Kennedy, W.M.; Sakaguchi, K.; Appella, E.;
              Gronenborn, A.M.
#journal        Nat. Struct. Biol. (1995) 2:321
#title          The solution structure of the oligomerization domain of
              the tumour suppressor p53.
REFERENCE      TN031005
#authors       Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;
              Sakamoto, H.; Appella, E.; Gronenborn, A.M.
#journal        Science (1995) 267:1515
#title          Interhelical angles in the solution structure of the
              oligomerization domain of p53: correction.
REFERENCE      TN031006
#authors       Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;
              Sakamoto, H.; Appella, E.; Gronenborn, A.M.
#journal        Science (1994) 265:386
#title          High-resolution structure of the oligomerization domain of
              p53 by multidimensional NMR.
COMMENT       Resolution: not applicable
COMMENT       Determination: NMR
KEYWORDS      anti-oncogene
FEATURE       17-37
              #region helix (right hand alpha)\
              #region beta sheet
SUMMARY       #length 42 #molecular-weight 4940 #checksum 6203

Query Match   100.0%: Score 78; DB 5; Length 42;
Best Local Similarity 100.0%: Pident. No. 1,18e-04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db            3 KPLDGEYFTL 12
ENTRY         |||||
TITLE         1 KPLDGEYFTL 10

RESULT        5
ENTRY         1SAHA      #type complete
TITLE         tumor suppressor p53 oligomerization domain, residues 319-360
              sac structures 2752, chain A - human
              high resolution solution NMR structure of the oligomerization
              domain of p53 by multi-dimensional NMR (sac structures)

```

|                            |  |
|----------------------------|--|
| ORGANISM                   | #domain of p53 by multi-dimensional NMR (sad structures)   |
| REFERENCE                  | #formal_name Homo sapiens #common_name man<br>A66595   |
| #authors                   | Clore, G.M.; Omichinski, J.G.; Gronenborn, A.M.  |
| #submission                | Submitted to The Brookhaven Protein Data Bank, March 1995  |
| #cross-references PDB:1SAH |  |
| REFERENCE                  | TN030986   |
| #authors                   | Clore, G.M.; Ernst, J.; Clubb, R.; Omichinski, J.G.;<br>Poundexter Kennedy, W.M.; Sakaguchi, K.; Appella, E.;<br>Gronenborn, A.M.                                      |
| REFERENCE                  |  |
| #journal                   | Nat. Struct. Biol. (1995) 2:321  |
| #title                     | Refined solution structure of the oligomerization domain of<br>the tumour suppressor p53.<br>TN030987  |
| #authors                   | Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;<br>Sakamoto, H.; Appella, E.; Gronenborn, A.M.   |
| REFERENCE                  |  |
| #journal                   | Science (1995) 267:1515  |
| #title                     | Interhelical angles in the solution structure of the<br>oligomerization domain of p53: correction.<br>TN030988   |
| #authors                   | Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;<br>Sakamoto, H.; Appella, E.; Gronenborn, A.M.   |
| COMMENT                    | High-resolution structure of the oligomerization domain of<br>p53 by multidimensional NMR.   |
| COMMENT                    | Resolution: not applicable   |
| COMMENT                    | Determination: NMR   |
| KEYWORDS                   | anti-Oncogene  |
| FEATURE                    |  |
| 17-37'                     |  |
| 9-15                       |  |
| SUMMARY                    |  |
| Query Match                | 100.0%; Score 78; DB 5; Length 42;   |
| Best Local Similarity      | 100.0%; Pred. No. 1.18e-04;  |
| Matches                    | 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  |
| Db                         | 3 KPDLGEFTL 12   |
|                            |  |
|                            | 1 KPDLGEFTL 10   |
| RESULT                     | 6  |
| ENTRY                      | 1SAHB #type complete   |
| TITLE                      | tumor suppressor p53 oligomerization domain, residues 319-360  |
| PDB_TITLE                  | sad structures 27-52, chain B - human  |
| ORGANISM                   | high resolution solution NMR structure of the oligomerization<br>domain of p53 by multi-dimensional NMR (sad structures)<br>#formal_name Homo sapiens #common_name man |
| REFERENCE                  | A66595   |
| #authors                   | Clore, G.M.; Omichinski, J.G.; Gronenborn, A.M.  |
| #submission                | submitted to the Brookhaven Protein Data Bank, March 1995  |
| #cross-references PDB:1SAH |  |
| REFERENCE                  | TN030989   |
| #authors                   | Clore, G.M.; Ernst, J.; Clubb, R.; Omichinski, J.G.;<br>Poundexter Kennedy, W.M.; Sakaguchi, K.; Appella, E.;<br>Gronenborn, A.M.                                      |
| REFERENCE                  |  |
| #journal                   | Nat. Struct. Biol. (1995) 2:321  |
| #title                     | Refined solution structure of the oligomerization domain of<br>the tumour suppressor p53.<br>TN030990  |
| #authors                   | Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;<br>Sakamoto, H.; Appella, E.; Gronenborn, A.M.   |
| REFERENCE                  |  |
| #journal                   | Science (1995) 267:1515  |
| #title                     | Interhelical angles in the solution structure of the<br>oligomerization domain of p53: correction.<br>TN030991   |
| #authors                   | Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;<br>Sakamoto, H.; Appella, E.; Gronenborn, A.M.   |
| REFERENCE                  |  |
| #journal                   | Science (1994) 265:386   |
| #title                     | High-resolution structure of the oligomerization domain of<br>p53 by multidimensional NMR.   |
| COMMENT                    | Resolution: not applicable   |

```

COMMENT      Determination: NMR
KEYWORDS     anti-oncogene

FEATURE
17-37        #region helix (right hand alpha)\
9-15         #region beta sheet
SUMMARY      #length 42 #molecular-weight 4940 #checksum 6203

Query Match      100.0%; Score 78; DB 5; Length 42;
Best Local Similarity 100.0%; Pred. No. 1.18e-04;
Matches          10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB      3      KPLDGEYFTL 12
        |||||
        |||||
QY      1      KPLDGEYFTL 10

RESULT      7
ENTRY      1SAHC      #type complete
TITLE      tumor suppressor p53 oligomerization domain, residues 319 360C
PDB_TITLE  sad structures 27 52, chain C - human
            high resolution solution NMR structure of the oligomerization
            domain of p53 by multi-dimensional NMR (sad structures)
ORGANISM   #formal_name Homo sapiens #common_name man
REFERENCE
#authors   Clore, G.M.; Omichinski, J.G.; Gronenborn, A.M.
#submission Submitted to the Brookhaven Protein Data Bank, March 1995
#cross-References PDB:1SAH
REFERENCE
#authors   Clore, G.M.; Ernst, J.; Clubb, R.; Omichinski, J.G.;
            Poldinger Kennedy, W.M.; Sakaguchi, K.; Appella, E.;
            Gronenborn, A.M.
            Net. Struct. Biol. (1995) 2:321
            Refined solution structure of the oligomerization domain of
            the tumour suppressor p53.
            TN030993
#journal   Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;
            Sakamoto, H.; Appella, E.; Gronenborn, A.M.
            Science (1995) 267:1515
            Interhelical angles in the solution structure of the
            oligomerization domain of p53: correction.
            TN030994
#authors   Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;
            Sakamoto, H.; Appella, E.; Gronenborn, A.M.
            Science (1994) 265:386
            High-resolution structure of the oligomerization domain of
            p53 by multidimensional NMR.
COMMENT     Resolution: not applicable
COMMENT     Determination: NMR
KEYWORDS    anti-oncogene

FEATURE
17-37        #region helix (right hand alpha)\
9-15         #region beta sheet
SUMMARY      #length 42 #molecular-weight 4940 #checksum 6203

Query Match      100.0%; Score 78; DB 5; Length 42;
Best Local Similarity 100.0%; Pred. No. 1.18e-04;
Matches          10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB      3      KPLDGEYFTL 12
        |||||
        |||||
QY      1      KPLDGEYFTL 10

RESULT      8
ENTRY      1SAHD      #type complete
TITLE      tumor suppressor p53 oligomerization domain, residues 319 360C
PDB_TITLE  sad structures 27 52, chain D - human
            high resolution solution NMR structure of the oligomerization
            domain of p53 by multi-dimensional NMR (sad structures)
ORGANISM   #formal_name Homo sapiens #common_name man
REFERENCE
#authors   Clore, G.M.; Omichinski, J.G.; Gronenborn, A.M.
#submission submitted to the Brookhaven Protein Data Bank, March 1995

```



|   |    |   |                     |            |    |
|---|----|---|---------------------|------------|----|
| Db  | 3  | KPLDGEYFTL 12   |                     |            |    |
| Qy  | 1  | KPLDGEYFTL 10   |                     |            |    |
| RESULT  | 13 |   |                     |            |    |
| ENTRY   |    | 1SAFC   | #type complete      |            |    |
| TITLE   |    | tumor suppressor p53 oligomerization domain, residues 319-360 |                     |            |    |
| PDB-TITLE   |    | sad structures 1 26, chain C - human                          |                     |            |    |
| ORGANISM  |    | high resolution solution NMR structure of the oligomerization |                     |            |    |
| REFERENCE   |    | domain of p53 by multi-dimensional NMR (sad structures)       |                     |            |    |
| #authors  |    | #formal_name Homo sapiens #common_name man                    |                     |            |    |
| #submission   |    | A66593  |                     |            |    |
| #cross-references   |    | Clore, G.M.; Omichinski, J.G.; Gronenborn, A.M.               |                     |            |    |
| REFERENCE   |    | submitted to the Brookhaven Protein Data Bank, March 1995     |                     |            |    |
| #authors  |    | TN030968  |                     |            |    |
| REFERENCE   |    | Clore, G.M.; Ernst, J.; Clubb, R.; Omichinski, J.G.;          |                     |            |    |
| #journal  |    | Poldekster Kennedy, W.M.; Sakaguchi, K.; Appella, E.;         |                     |            |    |
| #title  |    | Gronenborn, A.M.  |                     |            |    |
| REFERENCE   |    | Nat. Struct. Biol. (1995) 2:321                               |                     |            |    |
| #authors  |    | Refined solution structure of the oligomerization domain of   |                     |            |    |
| #journal  |    | the tumour suppressor p53.                                    |                     |            |    |
| #title  |    | TN030969  |                     |            |    |
| REFERENCE   |    | Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;   |                     |            |    |
| #journal  |    | Sakamoto, H.; Appella, E.; Gronenborn, A.M.                   |                     |            |    |
| #title  |    | Science (1995) 267:1515                                       |                     |            |    |
| REFERENCE   |    | Interhelical angles in the solution structure of the          |                     |            |    |
| #authors  |    | oligomerization domain of p53: correction.                    |                     |            |    |
| #journal  |    | TN030970  |                     |            |    |
| #title  |    | Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;   |                     |            |    |
| REFERENCE   |    | Sakamoto, H.; Appella, E.; Gronenborn, A.M.                   |                     |            |    |
| #journal  |    | Science (1994) 265:386  |                     |            |    |
| #title  |    | High-resolution structure of the oligomerization domain of    |                     |            |    |
| COMMENT   |    | p53 by multidimensional NMR.                                  |                     |            |    |
| COMMENT   |    | Resolution: not applicable                                    |                     |            |    |
| KEYWORDS  |    | Determination: NMR  |                     |            |    |
| FEATURE   |    | anti-oncogene   |                     |            |    |
| 17-37   |    |   |                     |            |    |
| 9-15  |    |   |                     |            |    |
| SUMMARY   |    | #region helix (right hand alpha)\                             |                     |            |    |
|   |    | #region beta sheet  |                     |            |    |
|   |    | #length 42 #molecular-weight 4940 #checksum 6203              |                     |            |    |
| Query Match   |    | 100.0%;   | Score 78;           | DB 5;      |    |
| Best Local Similarity                                       |    | 100.0%;   | Pred. No. 1,188-04; | Length 42; |    |
| Matches   |    | 10;   | Conservative        | 0;         |    |
|   |    | Mismatches  | 0;                  | Indels     | 0; |
|   |    | Gaps  | 0;                  |            |    |
| Db  | 3  | KPLDGEYFTL 12   |                     |            |    |
| Qy  | 1  | KPLDGEYFTL 10   |                     |            |    |
| RESULT  | 14 |   |                     |            |    |
| ENTRY   |    | 1SAFD   | #type complete      |            |    |
| TITLE   |    | tumor suppressor p53 oligomerization domain, residues 319-360 |                     |            |    |
| PDB-TITLE   |    | sad structures 1 26, chain D - human                          |                     |            |    |
| ORGANISM  |    | high resolution solution NMR structure of the oligomerization |                     |            |    |
| REFERENCE   |    | domain of p53 by multi-dimensional NMR (sad structures)       |                     |            |    |
| #authors  |    | #formal_name Homo sapiens #common_name man                    |                     |            |    |
| #submission   |    | A66593  |                     |            |    |
| #cross-references   |    | Clore, G.M.; Omichinski, J.G.; Gronenborn, A.M.               |                     |            |    |
| REFERENCE   |    | submitted to the Brookhaven Protein Data Bank, March 1995     |                     |            |    |
| #authors  |    | TN030971  |                     |            |    |
| REFERENCE   |    | Clore, G.M.; Ernst, J.; Clubb, R.; Omichinski, J.G.;          |                     |            |    |
| #journal  |    | Poldekster Kennedy, W.M.; Sakaguchi, K.; Appella, E.;         |                     |            |    |
| #title  |    | Gronenborn, A.M.  |                     |            |    |
| REFERENCE   |    | Nat. Struct. Biol. (1995) 2:321                               |                     |            |    |
| #journal  |    | Refined solution structure of the oligomerization domain of   |                     |            |    |
| #title  |    | the tumour suppressor p53.                                    |                     |            |    |
| TN030972  |    |   |                     |            |    |
| Clore, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.; |    |   |                     |            |    |

Search completed: Fri Sep 11 13:52:45 1998  
Job time : 31 secs.

```

#journal      Sakamoto, H.; Appella, E.; Gronenborn, A.M.
#title        Science (1995) 267:1515
#reference     Interhelical angles in the solution structure of the
#authors       oligomerization domain of p53: correction.
#journal      TN030973
#title        Clote, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;
#              Sakamoto, H.; Appella, E.; Gronenborn, A.M.
#              Science (1994) 265:386
#              High-resolution structure of the oligomerization domain of
#              p53 by multidimensional NMR.
#comment       Resolution: not applicable
#keywords      Determination: NMR
#feature       anti-oncogene
#length 42 #molecular-weight 4940 #checksum 6203
SUMMARY
  9-15      #region helix (right hand alpha)\
  17-37     #region beta sheet
  17-37     #region beta sheet
  9-15     #length 42 #molecular-weight 4940 #checksum 6203
SUMMARY
  Query Match 100.0%; Score 78; DB 5; Length 42;
  Best Local Similarity 100.0%; Pred. No. 1.18e-04;
  Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 KPLDGEYFTL 12
OY 1 KPLDGEYFTL 10

RESULT 15
ENTRY
TITLE      ISAF
PDB_TITLE  tumor suppressor p53 oligomerization domain, residues 319 360
            sad structures 1 26, chain A - human
            high resolution solution NMR structure of the oligomerization
            domain of p53 by multi-dimensional NMR (sad structures)
ORGANISM   #formal_name Homo sapiens #common_name man
REFERENCE  1A6593
#authors   Clote, G.M.; Omichinski, J.G.; Gronenborn, A.M.
#submission submitted to the Brookhaven Protein Data Bank, March 1995
#cross-references PDB:ISAF
REFERENCE  TN030962
#authors   Clote, G.M.; Ernst, J.; Clubb, R.; Omichinski, J.G.;
            Polindexter, Kennedy, W.M.; Sakaguchi, K.; Appella, E.;
            Gronenborn, A.M.
            Nat. Struct. Biol. (1995) 2:321
            Refined solution structure of the oligomerization domain of
            the tumour suppressor p53.
REFERENCE  TN030963
#authors   Clote, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;
            Sakamoto, H.; Appella, E.; Gronenborn, A.M.
            Science (1995) 267:1515
            Interhelical angles in the solution structure of the
            oligomerization domain of p53: correction.
REFERENCE  TN030964
#authors   Clote, G.M.; Omichinski, J.G.; Sakaguchi, K.; Zambrano, N.;
            Sakamoto, H.; Appella, E.; Gronenborn, A.M.
            Science (1994) 265:386
            High-resolution structure of the oligomerization domain of
            p53 by multidimensional NMR.
#comment     Resolution: not applicable
#keywords      Determination: NMR
#feature       anti-oncogene
#length 42 #molecular-weight 4940 #checksum 6203
SUMMARY
  Query Match 100.0%; Score 78; DB 5; Length 42;
  Best Local Similarity 100.0%; Pred. No. 1.18e-04;
  Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 KPLDGEYFTL 12
OY 1 KPLDGEYFTL 10

```





QY 1 KPLDGEYFTL 10

RESULT 2

ID P53-SHEEP STANDARD; PRT: 382 AA.

AC P51664;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53.

GN TP53.

CC 03 OVARY ARIES (SHEEP);

CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

CC EUTHERIA; ARTIODACTYLA.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-BLOOD;

RA MEDLINE; 95352828.

RA DEQUIEDT F., KETTMANN R., BURRY A., WILLEMS L.;

RL DNA SEQ. 5:255-259(1995).

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC EMBL: X81705; G602357; -

DR PROSITE: PS00348; P53; 1.

KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.

FT DOMAIN 1 66 ASP/GLU-RICH (ACIDIC).

FT MOD RES 300 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

FT MOD RES 381 381 PHOSPHORYLATION (BY SIMILARITY).

SO SEQUENCE 382 AA; 42809 MM; OCB99A00 CRC32;

Query Match 100.0%; Score 78; DB 1; Length 382;  
Best Local Similarity 100.0%; Pred. No. 6.48e-06;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 310 KPLDGEYFTL 319

QY 1 KPLDGEYFTL 10

RESULT 3

ID P53-FELCA STANDARD; PRT: 386 AA.

AC P41685;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53.

GN TP53.

CC FELIS SILVESTRIS CATUS (CAT);

CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

CC EUTHERIA; CARNIVORA.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-LYMPH NODE;

RA MEDLINE; 94333960.

RA OKUDA M., UMEIDA A., SAKAI T., OHASHI T., MOMOI Y., YOUN H.Y.,

RA MATSURI T., GOITSUKA R., TSUJIMOTO H., HASEGAWA A.;

RL INT. J. CANCER 58:602-607(1994).

RN [2]

RP SEQUENCE OF 34-354 FROM N.A.

RA MEDLINE; 94114699.

RA OKUDA M., UMEIDA A., MATSUMOTO Y., MOMOI Y., MATSURI T., GOITSUKA R.,

RA O'BRIEN S.J., TSUJIMOTO H., HASEGAWA A.;

RL J. VET. MED. SCI. 55:801-805(1993).

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.

CC EMBL: D26608; G538225; -

DR EMBL: D16460; G575528; -

DR PROSITE: PS00348; P53; 1.

KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.

FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).

FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

FT MOD RES 385 385 PHOSPHORYLATION (BY SIMILARITY).

FT CONFLICT 285 285 K -> R (IN REF. 2).

SO SEQUENCE 386 AA; 42692 MM; D6C7132A CRC32;

Query Match 100.0%; Score 78; DB 1; Length 386;  
Best Local Similarity 100.0%; Pred. No. 6.48e-06;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 314 KPLDGEYFTL 323

QY 1 KPLDGEYFTL 10

RESULT 4

ID P53-BOVIN STANDARD; PRT: 386 AA.

AC C29628;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53.

GN TP53.

CC BOS TAURUS (BOVINE);

CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

CC EUTHERIA; ARTIODACTYLA.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-LIVER;

RA MEDLINE; 95352829.

RA DEQUIEDT F., KETTMANN R., BURRY A., WILLEMS L.;

RL DNA SEQ. 5:261-264(1995).

RN [2]

RP SEQUENCE OF 13-386 FROM N.A.

RC STRAIN-HOLSTEIN; TISSUE-THYMUS;

RA MEDLINE; 96401400.

RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;

RL VET. IMMUNOL. IMMUNOPATHOL. 52:53-63(1996).

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.

1 -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: X81704; G602333;  
 DR EMBL: D49825; G1729419;  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOM RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 380 380 R -> T (IN REF. 2).  
 SQ SEQUENCE 386 AA; 43255 MW; 0322BF3D CRC32;  
 Query Match 100.0%; Score 78; DB 1; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 6,48e-06;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 314 KPLDGEYFTL 323  
 QY 1 KPLDGEYFTL 10  
 ID P53\_MOUSE STANDARD; PRT; 390 AA.  
 AC P02340;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR TRP53 OR P53.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 85027173.  
 RA BIENZ B., ZAKUT-HOURI R., GIOVOL D., OREN M.;  
 RL EMO J. 3; 2179-2183(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84068204.  
 RA ZAKUT-HOURI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIOVOL D.;  
 RL NATURE 306; 594-597(1983).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84272240.  
 RA JENKINS J.R., RUDGE K., REDMOND S., MADE-EVANS A.;  
 RL NUCLEIC ACIDS RES. 12; 5609-5626(1984).  
 RN [4]  
 RP SEQUENCE FROM N.A. (CLONES PCDS3; P53-M11 AND P53-M8).  
 RX MEDLINE: 87064640.  
 RA ARAI N., NOMURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.,  
 RL ROTHER V.;  
 RL MOL. CELL. BIOL. 6; 3232-3239(1986).  
 RN [5]  
 RP SEQUENCE OF 222-258 FROM N.A.  
 RX MEDLINE: 92115342.  
 RA BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BRENNER R.,  
 RL BALMAIN A.;  
 RL ONCOGENE 6; 2363-2369(1991).  
 RN [6]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE: 86149247.  
 RA SAMAD A., ANDERSON C.W., CARROLL R.B.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 83; 997-901(1986).  
 RN [7]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE: 91006019.  
 RA MEER D.W., SIMON S., KIKKAWA U., ECKHART W.;  
 RL EMO J. 9; 3253-3260(1990).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A

CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC EMBL: X00876; G871421; JOINED.  
 CC EMBL: X00877; G871421; JOINED.  
 CC EMBL: X00878; G871421; JOINED.  
 CC EMBL: X00879; G871421; JOINED.  
 CC EMBL: X00880; G871421; JOINED.  
 CC EMBL: X00881; G871421; JOINED.  
 CC EMBL: X00882; G871421; JOINED.  
 CC EMBL: X00883; G871421; JOINED.  
 CC EMBL: X00884; G871421; JOINED.  
 CC EMBL: X00885; G871421; JOINED.  
 CC EMBL: X01700; G200205; JOINED.  
 CC EMBL: X01237; G53576; JOINED.  
 CC EMBL: X00741; G53571; JOINED.  
 CC EMBL: M13872; G200199; JOINED.  
 CC EMBL: M13873; G200201; JOINED.  
 CC EMBL: M13874; G200203; ALT\_SEQ.  
 CC EMBL: S77930; G243255; JOINED.  
 CC PIR: A02684; DNMS53.  
 CC PIR: A22739; A22739.  
 CC PIR: S38822; S38822.  
 CC HSP: P04637; 1PES.  
 CC TRANSFAC: T01806; JOINED.  
 CC MGD: MGI:98834; TRP53.  
 CC PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS; DISEASE MUTATION.  
 FT DOMAIN 1 75 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 76 150 HYDROPHOBIC.  
 FT DOMAIN 276 390 HIGHLY BASIC AND MAY BE INVOLVED IN  
 FT INTERACTION WITH DNA.  
 FT DOMAIN 308 320 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 312 312 PHOSPHORYLATION.  
 FT MOD\_RES 389 389 PHOSPHORYLATION (BY CK2).  
 FT VARIANT 135 135 A -> V (CAN COOPERATE WITH AN ACTIVATED  
 FT RAS TO TRANSFORM FIBROBLASTS).  
 FT VARIANT 168 168 E -> G (IN CLONE P53-M11).  
 FT CONFLICT 48 48 O -> R (IN REF. 3).  
 FT CONFLICT 79 81 PVA -> QW (IN REF. 3).  
 SQ SEQUENCE 390 AA; 43458 MW; 8943DD93 CRC32;  
 Query Match 100.0%; Score 78; DB 1; Length 390;  
 Best Local Similarity 100.0%; Pred. No. 6,48e-06;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 318 KPLDGEYFTL 327  
 QY 1 KPLDGEYFTL 10  
 ID P53\_RAT STANDARD; PRT; 391 AA.  
 AC P10361; O09168;  
 DT 01-MAR-1989 (REL. 10, CREATED)  
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR P53.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE: 89083585.  
 RA SOUSST T.  
 RL NUCLEIC ACIDS RES. 16:11384-11384(1988).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 93181268.  
 RA HULLA J.E., SCHNEIDER R.P.  
 RL NUCLEIC ACIDS RES. 21:713-717(1993).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY.  
 RA MATHURPAIA S.P.  
 RL SUBMITTED (APR-1997) TO: EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC  
 CC EMBL: X13058; G56829; .  
 CC EMBL: L07910; G205952; JOINED.  
 CC EMBL: L07904; G205952; JOINED.  
 CC EMBL: L07905; G205952; JOINED.  
 CC EMBL: L07806; G205952; JOINED.  
 CC EMBL: L07907; G205952; JOINED.  
 CC EMBL: L07908; G205952; JOINED.  
 CC EMBL: L07909; G205952; JOINED.  
 CC EMBL: U90328; G1938365; .  
 CC PIR: S02192; S02192; .  
 CC HSSP: P04637; 1PRS; .  
 CC DR PROSITE: PS00348; P53; 1.  
 CC KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 CC NM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 CC FT DOMAIN 1 76 ASP/GLU-RICH (ACIDIC).  
 CC FT DOMAIN 77 151 HYDROPHOBIC.  
 CC FT DOMAIN 277 391 HIGHLY BASIC AND MAY BE INVOLVED IN  
 CC INTERACTION WITH DNA.  
 CC FT DOMAIN 309 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT MOD\_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).  
 CC FT VARIANT 103 103 G -> S.  
 CC FT VARIANT 256 256 E -> G.  
 CC FT CONFLICT 174 174 C -> W (IN REF. 2).  
 CC SQ SEQUENCE 391 AA; 43451 MW; ED114C18 CRC32;  
 CC  
 CC Query Match 100.0%; Score 78; DB 1; Length 391.  
 CC Best Local Similarity 100.0%; Pred. No. 6,488-06;  
 CC Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC Db 319 KPLDGEYFTL 328  
 CC 1 KPLDGEYFTL 10  
 CC  
 CC RESULT 7  
 CC ID P53\_HUMAN STANDARD; PRT; 393 AA.  
 CC AC P04637;  
 CC DT 13-AUG-1987 (REL. 05, CREATED)  
 CC DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
 CC DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 CC DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
 CC GN TP53.  
 CC OS HOMO SAPIENS (HUMAN).  
 CC OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 85230577.  
 RA ZAKUT-HOORI R., BIENZ-TADMOR B., GIVOL D., OREN M.;  
 RL EMBO J. 4:1251-1255(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87064416.  
 RA LAMB P., CRAWFORD L.;  
 RL MOL. CELL. BIOL. 6:1379-1385(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 85267676.  
 RA HATLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;  
 RL MOL. CELL. BIOL. 5:1601-1610(1985).  
 RN [4]  
 RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
 RX MEDLINE: 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTNER V.;  
 RL MOL. CELL. BIOL. 6:4650-4656(1986).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89108008.  
 RA BOCHMAN V.L., CHUMAKOV P.M., NIKKINA N.N., SAMARINA O.P.,  
 RA GEORGIEV G.P.;  
 RL GENE 70:245-252(1988).  
 RN [6]  
 RP SEQUENCE OF 101-393 FROM N.A.  
 RX MEDLINE: 85126934.  
 RA MATLASHENSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
 RA BENCHMOL S.;  
 RL EMBO J. 3:3257-3262(1984).  
 RN [7]  
 RP NUCLEAR LOCALIZATION SIGNAL.  
 RX MEDLINE: 90191730.  
 RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
 RL ONCOGENE 5:423-426(1990).  
 RN [8]  
 RP STRUCTURE BY NMR OF 319-360.  
 RX MEDLINE: 94294808.  
 RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
 RA APPELLEA E., GROENHOORN A.M.;  
 RL SCIENCE 265:386-391(1994).  
 RN [9]  
 RP STRUCTURE BY NMR OF 325-355.  
 RX MEDLINE: 95292092.  
 RA LEE W., HARVEY T.S., YIN Y., YANG P., LITCHFIELD D., ARROWSMITH C.H.;  
 RL NAT. STRUCT. BIOL. 1:877-890(1994).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
 RX MEDLINE: 94294806.  
 RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
 RL SCIENCE 265:346-355(1994).  
 RN [11]  
 RP REVIEW.  
 RX MEDLINE: 94090335.  
 RA HARRIS C.C.;  
 RL SCIENCE 262:1980-1981(1993).  
 RN [12]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE: 91289156.  
 RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
 RL SCIENCE 253:49-53(1991).  
 RN [13]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE: 96271983.  
 RA DE VRIES E.M.G., RICKS D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
 RA LIAO D., SOOSSI T., KOVACH J.S., SOMMER S.S.;  
 RL HUM. MUTAT. 7:202-213(1996).  
 RN [14]  
 RP VARIANT ARG-72.  
 RX MEDLINE: 91153807.

OLSCWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
 RL HUM. GENET. 86:369-370(1991).  
 RP VARIANT LFS THR-133.  
 RX MEDLINE: 92034774.  
 RA IAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
 RL CANCER RES. 51:6385-6387(1991).  
 RN [16]  
 RP VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
 RX MEDLINE: 91057657.  
 RA MAJLIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
 RA KIM H., KASSEL J., GRYKA M.A., BISCHOFF F.Z., TALINSKY M.A.,  
 RA FRIEND S.H.;  
 RL SCIENCE 250:1233-1236(1990).  
 RN [17]  
 RP VARIANT LFS ASP-245.  
 RX MEDLINE: 91080929.  
 RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
 RL NATURE 348:747-749(1990).  
 RN [18]  
 RP VARIANT LFS LEU-272.  
 RX MEDLINE: 9214783.  
 RA FELIX C.A., NAV M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
 RA POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
 RA KNUDSEN T., MINNA J.D.;  
 RL J. CLIN. INVEST. 89:640-647(1992).  
 RN [19]  
 RP VARIANTS LFS HIS-273 AND VAL-325.  
 RX MEDLINE: 92228023.  
 RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
 RA GEBHARDT M.C., ANDERSEN T.I., BORRESSEN A.-L., LI F.P., GARBER J.,  
 RA STRONG L.C.;  
 RL NEW ENGL. J. MED. 326:1309-1315(1992).  
 RN [20]  
 RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
 RX MEDLINE: 90295284.  
 RA BARBER J., IGGO R., GANNON J., LANE D.P.;  
 RL ONCOGENE 5:893-899(1990).  
 RN [21]  
 RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
 RX MEDLINE: 91017544.  
 RA RODRIGUES N.R., ROMAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
 RA GANNON J.V., LANE D.P.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 87:7555-7559(1990).  
 RN [22]  
 RP VARIANTS COLORECTAL CANCER MUTATIONS.  
 RX MEDLINE: 91282784.  
 RA ISHIOKA C., SATO T., GAMOH M., SUZUKI T., SHIBATA H., KANAMARU R.,  
 RA WAKUI A., YAMAZAKI T.;  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 177:901-906(1991).  
 RN [23]  
 RP VARIANTS OESOPHAGUS TUMORS L-152; A-155; H-175; F-176 AND H-273.  
 RX MEDLINE: 91330175.  
 RA CASSON A.G., MUKHOPADHYAY T., CLEARY K.R., RO J.Y., LEVIN B.,  
 RA ROTH J.A.;  
 RL CANCER RES. 51:4495-4499(1991).  
 RN [24]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN CHINA.  
 RX MEDLINE: 91187113.  
 RA HSU I.C., METCALF R.A., SUN T., WELSH J.A., WANG N.J., HARRIS C.C.;  
 RL NATURE 350:427-428(1991).  
 RN [25]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN SOUTH AFRICA.  
 RX MEDLINE: 91187114.  
 RA BRESSAC B., KEM M., WANDS J., OZTURK M.;  
 RL NATURE 350:429-431(1991).  
 RN [26]  
 RP VARIANTS IN ANAGENITAL CARCINOMAS.  
 RX MEDLINE: 93010989.  
 RA CROOK T., VOUSDEN K.H.;  
 RL EMBO J. 11:3935-3940(1992).  
 RN [27]  
 RP VARIANT WITH DUPLICATION OF PRO-177-HIS-178.

RX MEDLINE: 93265016.  
 RA BHATIA K., GUTTEREZ M.I., MAGRATH I.T.;  
 RL HUM. MOL. GENET. 1:207-208(1992).  
 RN [28]  
 RP VARIANTS IN BURKITT'S LYMPHOMAS.  
 RX MEDLINE: 93064692.  
 RA DUTRU A., DEBIRE B., ROMANO J.W., EHRHART J.C., FISCELLA M., MAY E.,  
 RA APPELLA E., MAY P.;  
 RL ONCOGENE 7:2161-2167(1992).  
 RN [29]  
 RP VARIANT NASOPHARYNGEAL CARCINOMA THR-280.  
 RX MEDLINE: 92335329.  
 RA SUN Y.-T., HEGAMER G., HENG Y.-J., HILDESHEIM A., CHEN J.-Y., CAO Y.,  
 RA YAO K.-T., COLBURN N.H.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:6516-6520(1992).  
 RN [30]  
 RP VARIANTS IN COLON TUMORS.  
 RX MEDLINE: 93330562.  
 RA HAMELIN R., IGGO N., LAURENT-PUIG P., VIDAUD M., THOMAS G.;  
 RL ONCOGENE 8:2213-2220(1993).  
 RN [31]  
 RP CHARACTERIZATION OF VARIANT ALA-143.  
 RX MEDLINE: 94283378.  
 RA ZHANG W., GUO X.-Y., HU G.-Y., LIU W.-B., SHAY J.W., DEISSEKOTH A.B.;  
 RL EMBO J. 13:2535-2544(1994).  
 RN [32]  
 RP VARIANTS LFS HIS-175; ARG-193; GLN-248; CYS-273 AND TYR-275.  
 RX MEDLINE: 95193787.  
 RA FREBOURG T., BARBIER N., VAN Y.-X., GARBER J.E., DREYFUS M.,  
 RA FRAUMENI J.F. JR., LI F.P., FRIEND S.H.;  
 RL AM. J. HUM. GENET. 56:608-615(1995).  
 RN [33]  
 RP VARIANT LFS HIS-175.  
 RX MEDLINE: 96423319.  
 RA VARLEY J.M., MCGROWN G., THORNCROFT M., TRICKER K.J., TEARE M.D.,  
 RA SANTIABANZ-KOREF M.F., HOUSTON R.S., MARTIN J., BIRCH J.M.,  
 RA EVANS D.G.R.;  
 RL J. MED. GENET. 32:942-945(1995).  
 RN [34]  
 RP VARIANTS BA PHE-176; SER-245; TRP-248; TRP-282 AND GLN-286.  
 RX MEDLINE: 96233927.  
 RA AUDREZET M.-P., ROBASKIEWICZ M., MERCIER B., NOUSBAUM J.-B.,  
 RA HARDY E., BAILL J.-P., VOLANT A., LOZAC'H P., GOUEROU H., FEREC C.;  
 RL HUM. MUTAT. 7:109-113(1996).  
 RN [35]  
 RP Note: remainder of annotations omitted.

Query Match 100.0%; Score 78; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6,48e-06;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 321 KPIDGEFTL 330  
 |||||  
 QY 1 KPIDGEFTL 10

RESULT 8  
 ID P53 CERAE STANDARD: PRT: 393 AA.  
 AC P13481.  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS CERCOPTHECUS AETHIOPS (GREEN MONKEY) (GRIYET).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE-LIVER;  
 RX MEDLINE: 90045967.  
 RA RIGAUDY P., ECKHARDT W.;  
 RL NUCLEIC ACIDS RES. 17:8375-8375(1989).

|                          |  |   |
|--------------------------|--|---|
| CC                       | -1-  | FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.                 |
| CC                       | -1-  | SUBCELLULAR LOCATION: NUCLEAR.  |
| CC                       | -1-  | DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.  |
| CC                       | -1-  | SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  |
| DR                       | EMBL:  | X16384; G22796; -   |
| DR                       | PIR:   | S06594; S06594.   |
| DR                       | HSSP:  | P04637; IOLG.   |
| DR                       | PROSITE:   | PS00348; P53. 1.  |
| KW                       | ANTI-ONCOGENE,   | DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;   |
| KW                       | NUCLEAR PROTEIN;   | PHOSPHORYLATION; APOPTOSIS.   |
| FT                       | DOMAIN   | 1 68 ASP/GLU-RICH (ACIDIC).   |
| FT                       | DOMAIN   | 81 150 HYDROPHOBIC.   |
| FT                       | DOMAIN   | 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN INTERACTION WITH DNA.   |
| FT                       | DOMAIN   | 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  |
| FT                       | MOD_RES  | 392 392 PHOSPHORYLATION (BY SIMILARITY).  |
| SO                       | SEQUENCE   | 393 AA; 43696 MW; BBETDC62 CRC32;   |
| Query Match              |  | 100.0%; Score 78; DR 1; Length 393;   |
| Best Local Similarity    |  | 100.0%; Pred. No. 6.48e-06;   |
| Matches 10; Conservative |  | 0; Mismatches 0; Indels 0; Gaps   |
| Dd                       | 321 KPDLGEYFLL 330   |   |
| Oy                       | 1 KPDLGEYFLL 10  |   |
| RESULT 9                 |  |   |
| ID                       | P53_RABIT  | STANDARD; PRT; 391 AA.  |
| AC                       | G95330;  |   |
| DT                       | 01-NOV-1997 (REL. 35, CREATED)   |   |
| DT                       | 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  |   |
| DT                       | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  |   |
| DE                       | CELLULAR TUMOR ANTIGEN P53.  |   |
| GN                       | TP53.  |   |
| OS                       | ORCHETOLOAGUS CUNICULUS (RABBIT).  |   |
| OC                       | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; EUTHERIA; LAGOMORPHA. |   |
| RN                       | [1]  |   |
| RP                       | SEQUENCE FROM N.A.   |   |
| RC                       | STRAIN-NEW ZEALAND;  |   |
| RX                       | MEDLINE: 97208869.   |   |
| RA                       | LE GOAS-F.; MAY P.; RONCO P.; CARON DE FROMENTEL C.;                                 |   |
| RL                       | GENE 185169-173(1997).   |   |
| CC                       | -1-  | FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION (BY SIMILARITY). |
| CC                       | -1-  | SUBCELLULAR LOCATION: NUCLEAR.  |
| CC                       | -1-  | DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.  |
| CC                       | -1-  | SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  |
| DR                       | EMBL:  | X90592; E194962; -  |
| DR                       | PROSITE:   | PS00348; P53. 1.  |
| DR                       | ANTI-ONCOGENE;   | DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;   |

|   |  |                                    |
|---|--|------------------------------------|
| KW  | NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.                         |                                    |
| FT  | DOMAIN 1 70 ASP/GLU-RICH (ACIDIC).                                   |                                    |
| FT  | DOMAIN 308 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).              |                                    |
| FT  | MOD_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).                     |                                    |
| SO  | SEQUENCE 391 AA; 43435 MM; 30A36172 CRC32;                           |                                    |
| Query Match 93.6%; Score 73; DB 1; Length 391;    |  |                                    |
| Best Local Similarity 90.0%; Pred. No. 1,12e-04;  |  |                                    |
| Matches   | 9; Conservative  | 0; Mismatches 1; Indels 0; Gaps 0  |
| Db  | 319 KPLDGEYFTL 328   |                                    |
| QY  | 1 KPLDGEYFTL 10  |                                    |
| RESULT 10   |  |                                    |
| ID  | P53_EQUUS  | STANDARD; PRT; 207 AA.             |
| AC  | Q29480:  |                                    |
| DT  | 01-NOV-1997 (REL. 35, CREATED)                                       |                                    |
| DT  | 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)                          |                                    |
| DT  | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)                        |                                    |
| DE  | CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).                               |                                    |
| GN  | P53.   |                                    |
| OS  | EQUUS ASINUS (DONKEY).   |                                    |
| OC  | EDUARROTA; METAOC; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;        |                                    |
| OC  | EUTHERIA; PERISSODACTYLA.  |                                    |
| RN  | [1]  |                                    |
| RP  | SEQUENCE FROM N.A.   |                                    |
| RX  | MEDLINE; 96342529.   |                                    |
| RA  | MASIR L., REID S.W.;   |                                    |
| RL  | DNA SEQ. 6:61-63(1995).  |                                    |
| CC  | -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES |                                    |
| CC  | GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL            |                                    |
| CC  | CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN      |                                    |
| CC  | TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A         |                                    |
| CC  | TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION   |                                    |
| CC  | BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF      |                                    |
| CC  | THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.     |                                    |
| CC  | APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF    |                                    |
| CC  | BAX AND BCL-2 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2          |                                    |
| CC  | EXPRESSION (BY SIMILARITY).  |                                    |
| CC  | -1- SUBCELLULAR LOCATION: NUCLEAR.                                   |                                    |
| CC  | -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY     |                                    |
| CC  | OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED       |                                    |
| CC  | IN MANY TYPES OF CANCER.   |                                    |
| CC  | -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.           |                                    |
| DR  | EMBL; U26414; G1020153; -  |                                    |
| DR  | PROSITE; PS00348; P53; 1.  |                                    |
| DR  | ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;     |                                    |
| KW  | NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.                         |                                    |
| RW  | NON_TER 1 1  |                                    |
| FT  | DOMAIN 187 199 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).              |                                    |
| FT  | NON_TER 207 207  |                                    |
| SO  | SEQUENCE 207 AA; 23428 MM; 0FBAB9C1 CRC32;                           |                                    |
| Query Match 91.0%; Score 71; DB 1; Length 207;    |  |                                    |
| Best Local Similarity 100.0%; Pred. No. 3,41e-04; |  |                                    |
| Matches   | 9; Conservative  | 0; Mismatches 0; Indels 0; Gaps 0; |
| Db  | 198 PLDGEYFTL 206  |                                    |
| QY  | 2 PLDGEYFTL 10   |                                    |
| RESULT 11   |  |                                    |
| ID  | P53_HORSE  | STANDARD; PRT; 280 AA.             |
| AC  | P79892; Q29481;  |                                    |
| DT  | 01-NOV-1997 (REL. 35, CREATED)                                       |                                    |
| DT  | 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)                          |                                    |
| DT  | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)                        |                                    |
| DE  | CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).                               |                                    |
| GN  | P53 OR P53.  |                                    |
| OS  | EQUUS CABALLUS (HORSE).  |                                    |

Sun Sep 13 10:56:25 1998

```

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PERISSODACTYLA.
RN [1].
RP SEQUENCE OF 1-263 FROM N.A.
RC TISSUE=SPLEEN;
RA MEDLINE: 97070350.
RL PAZZI K.A., KRAEGEL S.A., GRIFFEY S.M., THEON A.P., MADEWELL B.R.;
RN CANCER LETT. 107:125-130(1996).
[2]
RP SEQUENCE OF 76-280 FROM N.A.
RX MEDLINE: 96293865.
RL NASTR L., REID S.W.;
RL DNA SEQ. 6:185-187(1996).
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.
CC EMBL: S83123; G1836145; -
CC DR EMBL: U37120; G1389675; -
CC DR PROSITE: P500348; P53; 1.
CC KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
CC NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.
CC FT NON_TER 1 1
CC FT DOMAIN 262 274 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT CONFLICT 79 79 T -> A (IN REF. 2).
CC FT CONFLICT 83 83 L -> M (IN REF. 2).
CC FT CONFLICT 111 111 A -> V (IN REF. 2).
CC FT CONFLICT 138 138 G -> A (IN REF. 2).
CC FT NON_TER 280 280
CC FT SEQUENCE 280 AA; 30985 MW; B494F872 CRC32;
SQ
Query Match 89.7%; Score 70; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 5.92e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 272 KPLDGEYFT 280
QY 1 KPLDGEYFT 9

```

```

CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.
CC EMBL: Y08900; E303876; -
CC DR EMBL: Y08901; E303863; -
CC DR EMBL: U50395; G1842230; -
CC DR PROSITE: P500348; P53; 1.
CC KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
CC NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.
CC FT DOMAIN 1 74
CC FT DOMAIN 75 150 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 316 390 HIGHLY BASIC AND MAY BE INVOLVED IN
CC INTERACTION WITH DNA.
CC FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT MOD_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).
CC FT VARIANT 133 133 L -> Q (IN CELL LINE V79-4).
CC FT VARIANT 135 135 C -> W (IN CELL LINE V79-4).
CC FT CONFLICT 103 103 Y -> F (IN REF. 2).
CC FT SEQUENCE 393 AA; 43378 MW; 402EB149 CRC32;
SQ
Query Match 89.7%; Score 70; DB 1; Length 393;
Best Local Similarity 90.0%; Pred. No. 5.92e-04;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
DB 321 KPLDGEYFTL 330
QY 1 KPLDGEYFTL 10

```

```

RESULT 12
ID P53_CRIGR STANDARD; PRT; 393 AA.
AC 009185; 064397; P97258; P97788;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53.
DE TP53 OR P53.
OS CRICETUS GRISEUS (CHINESE HAMSTER).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1].
RP SEQUENCE FROM N.A.
RA CHANG W., MI L.J., BOORSTEIN R.J.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2].
RP SEQUENCE FROM N.A.
RX TISSUE=LIVER;
RX MEDLINE: 97183659.
RX LEE H., LARNER J.M., HAMLIN J.L.;
RX GENE 184:177-183(1997).
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES

```

```

RESULT 13
ID P53_MESAU STANDARD; PRT; 396 AA.
AC 000366; P97276;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53.
DE TP53.
OS MESOCRICEUS AURATUS (GOLDEN HAMSTER).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN-SYRIAN; TISSUE-KIDNEY;
RX MEDLINE: 92210007.
RX LEGROS Y., MCINTYRE P., SOUSSI T.;
RX GENE 112:247-250(1992).
RN [2].
RP SEQUENCE FROM N.A.
RA HOT E.W., WISEMAN R.;
RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY

```

OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: M75144; G191415; -;  
 DR EMBL: U07182; G473579; -;  
 DR PIR: JH0633; JH0633.  
 DR HSSP: P04637; JP0633.  
 DR PROSITE: PS00348; P53; 1.  
 DR ANTI-ONCOGENE: DNA-BINDING: TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 77 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 78 153 HYDROPHOBIC.  
 FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
 INTERACTION WITH DNA.  
 FT DOMAIN 314 326 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 395 395 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 188 188 G -> S (IN REF. 2).  
 SQ SEQUENCE 396 AA; 43631 MW; C2668ADE CRC32;  
 Query Match 89.7%; Score 70; DB 1; Length 396;  
 Best Local Similarity 90.0%; Pred. No. 5,92e-04;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Db 324 KTLDEYFTL 333  
 QY 1 KPLDEYFTL 10  
 RESULT 14  
 ID DHAS\_RAT STANDARD; PRT; 499 AA.  
 AC Q63639;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE RETINALDEHYDE-SPECIFIC DEHYDROGENASE TYPE 2 (EC 1.2.1.-) (RALDH(II))  
 DE (RALDH-2).  
 GN RALDH2.  
 OS RATUS NORVEGICUS (RAT).  
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1].  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TESTIS;  
 RX MEDLINE: 96279178.  
 RA WANG X.; PENZES P.; NAPOLI J.L.;  
 RL J. BIOL. CHEM. 271:16288-16293(1996).  
 CC -1- FUNCTION: RECOGNIZES AS SUBSTRATES FREE RETINAL AND CELLULAR  
 RETINOL-BINDING PROTEIN-BOUND RETINAL. DOES METABOLIZE OCTANAL AND  
 DECANAL BUT DOES NOT METABOLIZE CITRAL, BENZALDEHYDE, ACETALDEHYDE  
 AND PROPANAL EFFICIENTLY.  
 CC -1- PATHWAY: RETINOIC ACID BIOGENESIS.  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- TISSUE SPECIFICITY: FOUND IN TESTIS AND LESS ABUNDANTLY IN LUNG,  
 BRAIN, HEART, LIVER AND KIDNEY.  
 CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.  
 DR EMBL: U60063; G1403721; -;  
 DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; 1.  
 DR PROSITE: PS00687; ALDEHYDE\_DEHYDR\_GLU; 1.  
 KW OXIDOREDUCTASE; NAD.  
 FT NP\_BIND 244 249 NAD (ADP PART) (BY SIMILARITY).  
 FT ACT\_SITE 267 267 POTENTIAL.  
 FT ACT\_SITE 301 301 POTENTIAL.  
 SQ SEQUENCE 499 AA; 54739 MW; 8CF8E9E CRC32;  
 Query Match 71.8%; Score 56; DB 1; Length 499;  
 Best Local Similarity 66.7%; Pred. No. 8.39e-01;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 Db 144 PYDGDYFTL 152  
 QY 2 PLDGEYFTL 10

RESULT 15  
 ID DHAS\_MOUSE STANDARD; PRT; 499 AA.  
 AC Q62148;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE RETINALDEHYDE-SPECIFIC DEHYDROGENASE TYPE 2 (EC 1.2.1.-) (RALDH(II))  
 DE (RALDH-2).  
 GN RALDH2.  
 OS MUS MUSCULUS (MOUSE).  
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1].  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C3H/HE;  
 RX MEDLINE: 96390857.  
 RA ZHAO D.; MCCAFFERY P.; IVINS K.J.; NEVE R.L.; HOGAN P.; CHIN W.W.;  
 RA DRAEGER U.C.;  
 RL EUR. J. BIOCHEM. 240:15-22(1996).  
 CC -1- FUNCTION: RECOGNIZES AS SUBSTRATES FREE RETINAL AND CELLULAR  
 RETINOL-BINDING PROTEIN-BOUND RETINAL. DOES METABOLIZE OCTANAL AND  
 DECANAL BUT DOES NOT METABOLIZE CITRAL, BENZALDEHYDE, ACETALDEHYDE  
 AND PROPANAL EFFICIENTLY (BY SIMILARITY).  
 CC -1- PATHWAY: RETINOIC ACID BIOGENESIS.  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.  
 DR EMBL: X99273; E254167; -;  
 DR MGD: MGI:107928; RALDH2.  
 DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; 1.  
 DR PROSITE: PS00687; ALDEHYDE\_DEHYDR\_GLU; 1.  
 KW OXIDOREDUCTASE; NAD.  
 FT NP\_BIND 244 249 NAD (ADP PART) (BY SIMILARITY).  
 FT ACT\_SITE 267 267 POTENTIAL.  
 FT ACT\_SITE 301 301 POTENTIAL.  
 SQ SEQUENCE 499 AA; 54725 MW; 6B25ABBF CRC32;  
 Query Match 71.8%; Score 56; DB 1; Length 499;  
 Best Local Similarity 66.7%; Pred. No. 8.39e-01;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 Db 144 PYDGDYFTL 152  
 QY 2 PLDGEYFTL 10  
 Search completed: Fri Sep 11 13:53:10 1998  
 Job time : 7 secs.



# MIPS RELEASE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPearch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:53:27 1998; Maspar time 4.25 Seconds

Tabular output not generated. 99.161 Million cell updates/sec

Title: >US-08-452-843-20

Description: (1-10) from US08452843.pep

Perfect Score: 78

Sequence: 1 KPLDGEYFTL 10

Scoring table: PAM 150

Gap 15

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 08

Listing first 45 summaries

Database:

sptrembl6  
1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mmc 6:sp\_organelle 7:sp\_phage 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp\_vertebrate  
13:sp\_unclassified

Statistics: Mean 24.675; Variance 32.285; scale 0.764

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description            | Pred. No. |
|------------|-------|-------------|--------|----|--------|------------------------|-----------|
| 1          | 78    | 100.0       | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 2          | 78    | 100.0       | 381    | 13 | Q36005 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 3          | 78    | 100.0       | 393    | 2  | Q16811 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 4          | 78    | 100.0       | 393    | 2  | Q16807 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 5          | 78    | 100.0       | 393    | 2  | Q16807 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 6          | 78    | 100.0       | 393    | 2  | Q16848 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 7          | 78    | 100.0       | 393    | 2  | Q16810 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 8          | 78    | 100.0       | 393    | 2  | Q15086 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 9          | 78    | 100.0       | 393    | 2  | Q15086 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 10         | 78    | 100.0       | 393    | 2  | Q15088 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 11         | 78    | 100.0       | 393    | 2  | Q15087 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 12         | 78    | 100.0       | 393    | 2  | Q15087 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 13         | 78    | 100.0       | 393    | 2  | Q15087 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 14         | 78    | 100.0       | 393    | 2  | Q15087 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 15         | 78    | 100.0       | 393    | 2  | Q15087 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 16         | 78    | 100.0       | 393    | 2  | Q15087 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 17         | 78    | 100.0       | 393    | 2  | Q15087 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 18         | 78    | 100.0       | 393    | 2  | Q15087 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 19         | 78    | 100.0       | 393    | 2  | Q15087 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 20         | 78    | 100.0       | 393    | 2  | Q15087 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |

|    |    |      |      |    |        |                        |          |
|----|----|------|------|----|--------|------------------------|----------|
| 21 | 55 | 70.5 | 544  | 3  | Q20780 | SIMILAR TO ALDEHYDE DE | 3.06e+00 |
| 22 | 54 | 69.2 | 479  | 8  | Q42974 | NUCLEOTIDE PYROPHOSPHA | 4.83e+00 |
| 23 | 54 | 69.2 | 1379 | 3  | P91824 | T22A3.8.               | 4.83e+00 |
| 24 | 53 | 67.9 | 135  | 4  | Q29274 | UNKNOWN PROTEIN (FRAG  | 7.52e+00 |
| 25 | 53 | 67.9 | 238  | 9  | P74656 | HYPOHETICAL 26.6 KD P  | 7.52e+00 |
| 26 | 53 | 67.9 | 305  | 9  | Q58850 | HYPOHETICAL 35.0 KD P  | 7.52e+00 |
| 27 | 53 | 67.9 | 521  | 1  | Q04383 | D9719.7p.              | 7.52e+00 |
| 28 | 52 | 66.7 | 268  | 12 | Q91744 | XEL-A DNA-BINDING PROT | 1.16e+01 |
| 29 | 51 | 65.4 | 271  | 9  | Q28853 | RIBOSE-PHOSPHATE PYRO  | 1.79e+01 |
| 30 | 51 | 65.4 | 467  | 9  | Q32316 | HYPOHETICAL 52.1 KD P  | 1.79e+01 |
| 31 | 50 | 64.1 | 295  | 11 | Q83325 | MATRIX PROTEIN.        | 2.74e+01 |
| 32 | 50 | 64.1 | 300  | 11 | Q83324 | MATRIX PROTEIN.        | 2.74e+01 |
| 33 | 50 | 64.1 | 336  | 11 | Q83326 | ACUTE MEASLES VIRUS (S | 2.74e+01 |
| 34 | 50 | 64.1 | 336  | 11 | Q83327 | ACUTE MEASLES VIRUS (S | 2.74e+01 |
| 35 | 50 | 64.1 | 473  | 8  | Q40251 | VIOLAXANTHIN DE-EPOXID | 2.74e+01 |
| 36 | 49 | 62.8 | 232  | 8  | Q24261 | GLUTATHIONE S-TRANSFER | 4.17e+01 |
| 37 | 49 | 62.8 | 330  | 9  | Q55163 | HYPOHETICAL 35.9 KD P  | 4.17e+01 |
| 38 | 49 | 62.8 | 331  | 9  | Q30384 | PIIG.                  | 4.17e+01 |
| 39 | 49 | 62.8 | 351  | 9  | Q28599 | SN-GLYCEROL-1-PHOSPHAT | 4.17e+01 |
| 40 | 49 | 62.8 | 394  | 9  | Q06164 | MMGC.                  | 4.17e+01 |
| 41 | 49 | 62.8 | 946  | 1  | Q12369 | CHROMOSOME XII READING | 4.17e+01 |
| 42 | 49 | 62.8 | 1822 | 10 | Q35412 | SPA-1 LIKE PROTEIN P12 | 4.17e+01 |
| 43 | 48 | 61.5 | 101  | 9  | Q45108 | NITROGEN FIXATION PROT | 6.30e+01 |
| 44 | 48 | 61.5 | 197  | 3  | Q21465 | COSMID M0208.          | 6.30e+01 |
| 45 | 48 | 61.5 | 1785 | 3  | Q93781 | F53H4.1.               | 6.30e+01 |

## ALIGNMENTS

| RESULT | ID     | Query Match | Length | DB | ID     | Description            | Pred. No. |
|--------|--------|-------------|--------|----|--------|------------------------|-----------|
| 1      | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 2      | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 3      | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 4      | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 5      | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 6      | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 7      | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 8      | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 9      | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 10     | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 11     | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 12     | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 13     | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 14     | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 15     | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 16     | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 17     | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 18     | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 19     | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |
| 20     | Q29475 | 100.0%      | 281    | 4  | Q29475 | CELLULAR TUMOR ANTIGEN | 3.06e-05  |

AC 036006:1  
 DT 01-JAN-1998 (TREMELREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMELREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN P53.  
 OS MARMATA MONAX.  
 OG PLASMID PT7BLUE (R).  
 CC UNCLASSIFIED.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA FELTELSON N.A., RANGANATHAN P.N., CLAYTON M.M., ZHANG S.M.;  
 RL ONCOGENE 15:327-336(1997).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: AJ001022; E351287; -.  
 DR PROSITE: PS00346; P53; 1.  
 KW PLASMID; ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION;  
 KM ACTIVATOR; NUCLEAR PROTEIN; PHOSPHORYLATION.  
 SO SEQUENCE 391 AA; 43468 MW; 95FAB8F2 CRC32;  
 Query Match 100.0%; Score 78; DB 13; Length 391;  
 Best Local Similarity 100.0%; Pred. No. 3.06e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 319 KPLDGEYFTL 328  
 ||||||||  
 QY 1 KPLDGEYFTL 10  
 RESULT 3  
 ID 016811 PRELIMINARY; PRT; 393 AA.  
 AC 016811:  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMELREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 85126934.  
 RA MATLSHEWSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
 RA BENCHIMOL S.;  
 RL EMO J. 3:3257-3262(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87064416.  
 RA LAMB P., CRAWFORD L.;  
 RL MOL. CELL. BIOL. 6:1379-1385(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: M13121; G386994; -.  
 DR EMBL: M13112; G386994; JOINED.  
 DR EMBL: M13113; G386994; JOINED.  
 DR EMBL: M13114; G386994; JOINED.  
 DR EMBL: M13115; G386994; JOINED.  
 DR EMBL: M13116; G386994; JOINED.  
 DR EMBL: M13117; G386994; JOINED.  
 DR EMBL: M13118; G386994; JOINED.  
 DR EMBL: M13119; G386994; JOINED.  
 DR EMBL: M13120; G386994; JOINED.

DR PROSITE: PS00346; P53; 1.  
 KW REPEAT; TUMOR ANTIGEN; ANTI-ONCOGENE; DNA-BINDING;  
 KM TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN;  
 KW PHOSPHORYLATION.  
 FT NON\_TER 393 393  
 SO SEQUENCE 393 AA; 43698 MW; 3EA71431 CRC32;  
 Query Match 100.0%; Score 78; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.06e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 KPLDGEYFTL 330  
 ||||||||  
 QY 1 KPLDGEYFTL 10  
 RESULT 4  
 ID 016807 PRELIMINARY; PRT; 393 AA.  
 AC 016807:  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMELREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: X60011; G506435; -.  
 DR PROSITE: PS00346; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 193 193 R -> H.  
 FT NON\_TER 393 393  
 SO SEQUENCE 393 AA; 43731 MW; 279BC9CB CRC32;  
 Query Match 100.0%; Score 78; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.06e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 KPLDGEYFTL 330  
 ||||||||  
 QY 1 KPLDGEYFTL 10  
 RESULT 5  
 ID 016808 PRELIMINARY; PRT; 393 AA.  
 AC 016808:  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMELREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT

CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: X60018; G506449; -  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 163 163 H -> Y.  
 FT NON\_TER 393 393  
 SO SEQUENCE 393 AA; 43627 MW; AFD8A9E3 CRC32;  
 Query Match 100.0%; Score 78; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.06e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 KPLDGEYFTL 330  
 QY 1 KPLDGEYFTL 10  
 RESULT 6 PRELIMINARY; PRT; 393 AA.  
 ID Q16848;  
 AC Q16848;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE; 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA RORTER V.;  
 RA MOL. CELL. BIOL. 6:4650-4656(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: M14694; G339814; -  
 DR PROSITE: PS00348; P53; 1.  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ANTI-ONCOGENE; DNA-BINDING;  
 KW TRANSCRIPTION REGULATION; ACTIVATOR.  
 SO SEQUENCE 393 AA; 43723 MW; DA7D302F CRC32;  
 Query Match 100.0%; Score 78; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.06e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 KPLDGEYFTL 330  
 QY 1 KPLDGEYFTL 10  
 RESULT 7 PRELIMINARY; PRT; 393 AA.  
 ID Q16810;  
 AC Q16810;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: X60020; G506453; -  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 254 254 D -> N.  
 FT VARIANT 254 254 D -> V.  
 FT NON\_TER 393 393  
 SO SEQUENCE 393 AA; 43714 MW; 5F914579 CRC32;  
 Query Match 100.0%; Score 78; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.06e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 KPLDGEYFTL 330  
 QY 1 KPLDGEYFTL 10  
 RESULT 8 PRELIMINARY; PRT; 393 AA.  
 ID Q15086;  
 AC Q15086;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE; 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL: X60013; G506439; -  
 FT VARIANT 246 246 T -> M.  
 FT NON\_TER 393 393  
 SO SEQUENCE 393 AA; 43682 MW; 943B62A3 CRC32;  
 Query Match 100.0%; Score 78; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.06e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 KPLDGEYFTL 330  
 QY 1 KPLDGEYFTL 10  
 RESULT 9 PRELIMINARY; PRT; 393 AA.  
 ID Q16535;  
 AC Q16535;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL: X60017; G506447; -.  
 DR EMBL: X60015; G506443; -.  
 FT VARIANT 248 248 Q -> R.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43684 MW; 239818A9 CRC32;  
 Query Match 100.0%; Score 78; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.06e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 KPLDGEFTL 330  
 QY 1 KPLDGEFTL 10

RESULT 10  
 ID 015088 PRELIMINARY; PRT; 393 AA.  
 AC 015088;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL: X60016; G506445; -.  
 FT VARIANT 238 238 Y -> C.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43713 MW; A01E1523 CRC32;  
 Query Match 100.0%; Score 78; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.06e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 KPLDGEFTL 330  
 QY 1 KPLDGEFTL 10

RESULT 11  
 ID 016809 PRELIMINARY; PRT; 393 AA.  
 AC 016809;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DE 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: X60019; G506451; -.  
 FT PROSITE: PS00348; P53; 1.

KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 213 213 Q -> R.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;  
 Query Match 100.0%; Score 78; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.06e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 KPLDGEFTL 330  
 QY 1 KPLDGEFTL 10

RESULT 12  
 ID 015087 PRELIMINARY; PRT; 393 AA.  
 AC 015087;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL: X60014; G506441; -.  
 FT VARIANT 237 237 I -> M.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43694 MW; 9BB81992 CRC32;  
 Query Match 100.0%; Score 78; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.06e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 KPLDGEFTL 330  
 QY 1 KPLDGEFTL 10

RESULT 13  
 ID 029484 PRELIMINARY; PRT; 196 AA.  
 AC 029484;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS EQUUS CABALLUS (HORSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PERISSODACTYLA.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA BUCHER K., SZALAI G., MARTI E., PAULI U., LAZARY S.;  
 RL RES. VET. SCI. 0:0-0(0).  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: X91793; E218035; -.  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT NON\_TER 196 196

SEQUENCE : 196 AA; 22080 MW; F443239C CRC32;

|                       |         |           |       |             |
|-----------------------|---------|-----------|-------|-------------|
| Query/Match           | 89.78;  | Score 70; | DB 4; | Length 196; |
| Best/Total Similarity | 100.08; | Prod VC   | 3     | 100-03.     |

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 188 KPLDGEYFT 196  
QY 1 KPLDGEYFT 9

RESULT 14  
ID 035873  
PRELIMINARY; PRT; 205 AA

0358/3;  
01-JAN-1998 (TREMORIET. 05-CREATED)

|    |             |             |                       |                         |
|----|-------------|-------------|-----------------------|-------------------------|
|    | (TREMBLREL. | .05,        | LAST SEQUENCE UPDATE) |                         |
| DT | 01-JAN-1998 | (TREMBLREL. | .05,                  | LAST ANNOTATION UPDATE) |
| DT | 01-JAN-1998 | (TREMBLREL. | .05,                  | LAST ANNOTATION UPDATE) |

DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT)  
GN P53.

OS CRICETULUS GRISEUS (CHINESE HAMSTER).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.

|    |                     |
|----|---------------------|
| OC | EUTHERIA; RODENTIA. |
| RN | [1]                 |
| DD | STOMYX              |

RA RAINALDI G., MARCHETTI S., CAPECCHI B., MENEVERI R., PIRAS A.,  
I FIZIO D.

DECEMBER 1997  
 SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS  
 RN [2]

RP SEQUENCE FROM N.A.  
RA VATTERONI L., MUSIO A., MENEVERI R., RAINALDI G.;

RL SUBMITTED (NOV-1997) TO EMBL/GENEBANK/DBJ DATA BANKS  
CC -I- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR

CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY IT  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT AC

CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS, ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC GROWTH AND DIVISION, THAT IS, THE GENES ACTIVATED ARE

CC - Cyclicin-Dependent Kinases (by  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DB - EMBL: U74487; G3581764; -

DR PROSITE: PS00348; P53; 1.  
KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTON REGULATION; ACTIVATOR

| Accession | Protein                           | Phosphorylation |
|-----------|-----------------------------------|-----------------|
| KW        | NUCLEAR PROTEIN; PHOSPHORYLATION. |                 |
| FT        | NON_TER 1                         | 1               |

|    |          |         |                           |
|----|----------|---------|---------------------------|
| FT | NON_TER  | 205     | 205                       |
| SQ | SEQUENCE | 205 AA; | 23122 MW; 680DDDDDC CRC32 |

| Query Match | 89.7% | Score 70 | DB 10 | Length 205 |
|-------------|-------|----------|-------|------------|
|-------------|-------|----------|-------|------------|

|                       |       |              |           |
|-----------------------|-------|--------------|-----------|
| Best Local Similarity | 90.0% | Pred. No.    | 2.10e-03; |
| Matches               | 9;    | Conservative | 0;        |
|                       |       | Mismatches   | 1;        |
|                       |       | Indels       | 0;        |
|                       |       | Gaps         | 0         |

Db 196 KTLDDGEYFTL 205

QY 1 KPLDGEYFTL 10

## RESULT 15

ID P89004 PRELIMINARY; PRT; 238 AA.  
 AC P89004;

| DT | 01-MAY-1997 | (TREMBLREL. 03, CREATED)              |
|----|-------------|---------------------------------------|
| DT | 01-MAY-1997 | (TREMBLREL. 03, LAST SEQUENCE UPDATE) |
| DT | 01-MAY-1997 | (TREMBLREL. 03, LAST SEQUENCE UPDATE) |

DE P53 (FRAGMENT):  
OS MASTOMYS NATALENSIS PAPILLOMAVIRUS (MURDER)  
DI 01 MAY 1991 (IREMBREL. 03, LAST ANNOTATION UPDATE

OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPILOMAVIRUSES  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=ECLOMA INDUCED BY LOXTIDINE.

RA LUQUE E.A., TANG L.H., MODLIN I.M.;  
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS

|    |                            |
|----|----------------------------|
| DR | EMBL; U48618; G1813455; -. |
| FT | NON_TER 1 1                |
| CO | SEQUENCE 330 1 570 1       |

SEQUENCE 238 AA; 26/04 MW; 09/E01F9 CRC32

| Query Match           | Score | DB | Length |
|-----------------------|-------|----|--------|
| Best Local Similarity | 88.58 | 69 | 238    |
| Best Local Similarity | 80.08 | 2  | 51     |

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

214 KKLDGEIFTL 223

QY 1 KPLDGEYFTL 10

Search completed: Fri Sep 11 13:54:03 1998  
Job time : 36 secs.

**THIS PAGE BLANK (USPTO)**

# W22024

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 13 10:41:29 1998; Maspar time 2.83 Seconds  
51.478 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-19  
Description: (1-9) from US08452843.pep  
Perfect Score: 64  
Sequence: 1 SPQPKKKPL 9

Scoring table:  
PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

a-geneseg32  
1:part1 2:part2 3:part3 4:part4 5:parts 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 15.081; Variance 44.744; scale 0.337  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length DB | ID     | Description           | Pred. No. |
|------------|-------|-------------|-----------|--------|-----------------------|-----------|
| 1          | 64    | 100.0       | 73 26     | W22024 | Wild-type p53 tetrame | 7.06e-01  |
| 2          | 64    | 100.0       | 319 24    | W28496 | Human p53 protein var | 7.06e-01  |
| 3          | 64    | 100.0       | 319 24    | W28495 | Human p53 protein var | 7.06e-01  |
| 4          | 64    | 100.0       | 335 24    | W28498 | Human p53 protein var | 7.06e-01  |
| 5          | 64    | 100.0       | 335 24    | W28497 | Human p53 protein var | 7.06e-01  |
| 6          | 64    | 100.0       | 353 24    | W28494 | Human p53 protein var | 7.06e-01  |
| 7          | 64    | 100.0       | 353 24    | W28493 | Human p53 protein var | 7.06e-01  |
| 8          | 64    | 100.0       | 355 22    | W13950 | Human p53 protein var | 7.06e-01  |
| 9          | 64    | 100.0       | 351 21    | W13958 | Chimeric p53 protein  | 7.06e-01  |
| 10         | 64    | 100.0       | 363 24    | W28480 | Human p53 protein var | 7.06e-01  |
| 11         | 64    | 100.0       | 363 24    | W13971 | Human p53 protein var | 7.06e-01  |
| 12         | 64    | 100.0       | 363 21    | W13974 | Modified p53 variant  | 7.06e-01  |
| 13         | 64    | 100.0       | 363 21    | W13973 | Modified p53 variant  | 7.06e-01  |
| 14         | 64    | 100.0       | 363 21    | W13972 | Modified p53 variant  | 7.06e-01  |
| 15         | 64    | 100.0       | 363 21    | W13975 | Modified p53 variant  | 7.06e-01  |
| 16         | 64    | 100.0       | 363 21    | W13977 | Modified p53 variant  | 7.06e-01  |
| 17         | 64    | 100.0       | 370 21    | W13957 | Chimeric p53 protein  | 7.06e-01  |
| 18         | 64    | 100.0       | 370 21    | W13957 | Chimeric p53 protein  | 7.06e-01  |

|    |    |       |        |        |                       |          |
|----|----|-------|--------|--------|-----------------------|----------|
| 19 | 64 | 100.0 | 374 24 | W28482 | Human p53 protein var | 7.06e-01 |
| 20 | 64 | 100.0 | 374 24 | W28481 | Human p53 protein var | 7.06e-01 |
| 21 | 64 | 100.0 | 381 24 | W28489 | Human p53 protein var | 7.06e-01 |
| 22 | 64 | 100.0 | 381 24 | W28490 | Human p53 protein var | 7.06e-01 |
| 23 | 64 | 100.0 | 393 22 | W25155 | Human p53 variant fou | 7.06e-01 |
| 24 | 64 | 100.0 | 393 22 | W13953 | Human modified human  | 7.06e-01 |
| 25 | 64 | 100.0 | 393 22 | W13948 | Human wild-type p53 t | 7.06e-01 |
| 26 | 64 | 100.0 | 393 21 | W05345 | Human p53 mutant N239 | 7.06e-01 |
| 27 | 64 | 100.0 | 393 22 | W13951 | Human tumour-derived  | 7.06e-01 |
| 28 | 64 | 100.0 | 393 22 | W13949 | Human tumour-derived  | 7.06e-01 |
| 29 | 64 | 100.0 | 393 22 | W13979 | Human tumour-derived  | 7.06e-01 |
| 30 | 64 | 100.0 | 393 22 | W13952 | Human tumour-derived  | 7.06e-01 |
| 31 | 64 | 100.0 | 393 19 | W02617 | Human p53 tumour supp | 7.06e-01 |
| 32 | 64 | 100.0 | 393 18 | R91933 | Wild type p53 protein | 7.06e-01 |
| 33 | 64 | 100.0 | 393 21 | W05348 | Human p53 mutant R282 | 7.06e-01 |
| 34 | 64 | 100.0 | 393 21 | W05344 | Human p53             | 7.06e-01 |
| 35 | 64 | 100.0 | 393 21 | W13970 | Modified p53 variant  | 7.06e-01 |
| 36 | 64 | 100.0 | 393 21 | W13968 | Modified p53 variant  | 7.06e-01 |
| 37 | 64 | 100.0 | 401 24 | W28487 | Human p53 protein var | 7.06e-01 |
| 38 | 64 | 100.0 | 401 24 | W28488 | Human p53 protein var | 7.06e-01 |
| 39 | 64 | 100.0 | 402 21 | W13965 | Chimeric p53 protein  | 7.06e-01 |
| 40 | 64 | 100.0 | 404 21 | W13963 | Chimeric p53 protein  | 7.06e-01 |
| 41 | 64 | 100.0 | 406 21 | W13966 | Chimeric p53 protein  | 7.06e-01 |
| 42 | 64 | 100.0 | 411 21 | W13967 | Chimeric p53 protein  | 7.06e-01 |
| 43 | 64 | 100.0 | 533 23 | W19763 | p53-GM-CSF immunostim | 7.06e-01 |
| 44 | 64 | 100.0 | 535 24 | W28492 | Human p53 protein var | 7.06e-01 |
| 45 | 64 | 100.0 | 535 24 | W28491 | Human p53 protein var | 7.06e-01 |

## ALIGNMENTS

RESULT 1  
ID W22024 standard; Protein: 73 AA.  
AC W22024;  
DR 09-MAR-1998 (first entry)  
DE Wild-type p53 tetramerising domain.  
KW Globin analogue; GCNA; yeast transcription factor; oligomerising domain;  
KW ligand binding domain; multimeric haemoglobin; oxygen carrier; anaemia;  
KW blood substitute; therapy; haematopoiesis; oxygen removal; HB; p53;  
OS nitric oxide removal; tetramerising domain.  
PN W09723631-R2.  
PD 03-JUL-1997.  
PF 20-DEC-1996; U20632.  
PR 22-DEC-1995; US-021001.  
PA (SOMA-) SOMATOGEN INC.  
PI Anthony-Cahill SJ, Epp JK, Kerwin BA, Mathews AJ;  
PI Olin PO;  
DR WPT 97-351067/32.  
PT New globin containing non-natural binding site and related nucleic  
PT acid - also multimeric haemoglobin, used as oxygen carrier for in  
PT vivo or in vitro applications, with extended half-life and reduced  
PT extravasation  
PS Example 8: Page 40; 64pp; English.  
CC This sequence represents the tetramerising domain of p53. This sequence,  
CC or the oligomerising domains of the yeast transcription factor GCNA (see  
CC W22019 and W22020) can be used in the globin of the invention. The  
CC globin of the invention has a non-natural binding domain (BD), preferably  
CC an oligomerising domain or ligand binding domain. The globin may be  
CC combined with other globins to form a multimeric haemoglobin (HB). The HB  
CC are used as oxygen carriers, both in vivo (as blood substitutes, volume  
CC extenders, in treatment of anaemia and to stimulate haematopoiesis) and  
CC in vitro (e.g. to improve growth of cell cultures). They are also used to  
CC remove oxygen from solutions, or therapeutically to remove nitric oxide.  
CC The HB can also be used as a reference standard for analytical  
CC instruments and for delivering drugs or in in vivo imaging. Incorporation  
CC of the BD allows production of larger HB that can be assembled without  
CC using exogenous crosslinking agents, and the size of the multimer can be  
CC controlled. Large HB show reduced extravasation and prolonged half life,  
CC and are able to deliver oxygen to tissues which erythrocytes can not  
CC reach (e.g. downstream of a thrombus, angioplasty balloon etc.).  
CC Sequence 73 AA.

Query Match 100.0%; Score 64; DB 26; Length 73;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 21 spgpkkpl 29  
 |||||||  
 1 SPQPKKPL 9

RESULT 2  
 ID W28496 standard; Protein: 319 AA.  
 AC W28496;  
 DT 25-NOV-1997 (first entry);  
 DE Human p53 protein variant 360-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis;  
 OS Homo sapiens.  
 OS Synthetic.

FT misc-difference 145 Location/Qualifiers  
 FT /note= "Arg residue at position 182 of wild-type  
 FT p53 has been mutated to His"  
 PD 06-FEB-1997.  
 PR 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 38; Page -; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360-325H and comprising  
 CC the 325-360 domain, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 360-325).  
 SQ Sequence 319 AA;

Query Match 100.0%; Score 64; DB 24; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 279 spgpkkpl 287  
 |||||||  
 1 SPQPKKPL 9

PN W09704092-A1.  
 PD 06-FEB-1997.  
 PR 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB: 786223.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 38; Pages 92-94; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360-325 and comprising  
 CC the 325-360 domain, amino acids 75-325 of human wild-type p53 and a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 319 AA;

Query Match 100.0%; Score 64; DB 24; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 279 spgpkkpl 287  
 |||||||  
 1 SPQPKKPL 9

RESULT 4  
 ID W28498 standard; Protein: 335 AA.  
 AC W28498;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360h-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; hinge region;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis;  
 OS Homo sapiens.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT region 39..53 /label= hinge  
 FT misc-difference 161  
 FT /note= "Arg residue at position 182 of wild-type  
 FT p53 has been mutated to His"  
 PD 06-FEB-1997.  
 PR 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 39; Page -; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360h-325H and comprising  
 CC the 325-360 domain, separated from amino acids 75-325 of human  
 CC wild-type p53 (but with Arg182 replaced by His) by a synthetic hinge  
 CC sequence (GlySer)3, and with a leucine zipper domain at the C-terminal.



CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 360h-325).  
 SQ Sequence 335 AA;

Query Match 100.0%; Score 64; DB 24; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 295 spqpkkkp1 303  
 |||||||  
 QY 1 SPQPKKKPL 9

RESULT 5  
 ID W28497 standard; Protein: 335 AA.  
 AC W28497;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360h-325 encoded by pEC179.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutetin;  
 KW substitution; replacement; transactivation; hinge region;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 FH Synthetic.  
 FT Key  
 FT region Location/Qualifiers  
 FT 39..53  
 FT /label= hinge  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PE 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB: T86224.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 37; Pages 94-95; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360h-325 and comprising  
 CC the 325-360 domain, separated from amino acids 75-325 of human  
 CC wild-type p53 by a synthetic hinge sequence (Gly4Ser)3, and with a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 335 AA;

Query Match 100.0%; Score 64; DB 24; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 295 spqpkkkp1 303  
 |||||||  
 QY 1 SPQPKKKPL 9

RESULT 6  
 ID W28494 standard; Protein: 353 AA.  
 AC W28494;

DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 393-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutetin;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 FH Synthetic.  
 FT Key  
 FT misc\_difference Location/Qualifiers  
 FT 179  
 FT /note= "Arg residue at position 182 of wild-type  
 FT p53 has been mutated to His"  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PE 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 37; Page -: 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-393 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 393-325H and comprising  
 CC the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 393-325).  
 SQ Sequence 353 AA;

Query Match 100.0%; Score 64; DB 24; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 313 spqpkkkp1 321  
 |||||||  
 QY 1 SPQPKKKPL 9

RESULT 7  
 ID W28493 standard; Protein: 353 AA.  
 AC W28493;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 393-325 encoded by pEC177.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutetin;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 FH Synthetic.  
 FT Key  
 FT PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PE 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB: T86222.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 37; Pages 90-92; 133pp; French.

CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-393 of p53. The present sequence is that of  
 CC a specifically claimed p53 variant designated 393-325 and comprising  
 CC the 325-393 domain, amino acids 75-325 of human wild-type p53 and a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SO Sequence 353 AA;

Query Match 100.0%; Score 64; DB 24; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 313 spqpkkkp1 321  
 Oy 1 SPOPKKKPL 9

RESULT 8  
 ID W13950 standard; Protein; 355 AA.  
 AC W13950;  
 DT 25-JUN-1997 (first entry)  
 DE Del356-393 modified human p53.  
 KM p53; tumour suppressor; cancer; therapy: cell proliferation;  
 KM apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN W09710843-A1.  
 PD 27-MAR-1997.  
 PE 20-SEP-1996; U5188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazoneis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer.  
 PS Claim 3; Refer to Page 27-29; 82pp; English.  
 CC Del356-393 modified p53 (W13950) has the C-terminal region of  
 CC wild-type human p53 tumour suppressor (W13948) deleted. Modified  
 CC p53 constructs (see also W13954, W13956-61, W13971-77) bearing  
 CC a deletion of all or a fragment of the C-terminal residues  
 CC 356-393 have DNA binding ability and can activate the DNA binding  
 CC of common class I p53 tumour mutants (see also W13951-52). The  
 CC method provides the means for pharmacological rescue of p53  
 CC function in cancer patients. Nucleic acids coding for such  
 CC constructs can be used for cancer gene therapy.  
 SO Sequence 355 AA;

Query Match 100.0%; Score 64; DB 22; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 spqpkkkp1 323  
 Oy 1 SPOPKKKPL 9

RESULT 9  
 ID W13958 standard; Protein; 361 AA.  
 AC W13958;  
 DT 25-JUN-1997 (first entry)  
 DE Chimeric p53 protein.  
 KM p53; tumour suppressor; cancer; therapy: cell proliferation;  
 KM apoptosis; protein engineering; GCN4; DNA binding.  
 OS Chimeric Homo sapiens;  
 OS Chimeric synthetic.  
 FH Key Location/Qualifiers

FT region 1..325  
 FT /label="p53wt  
 FT /note="amino acids 1-325 of wild-type p53"  
 FT region 326..328  
 FT /label="Linker  
 FT region 329..361  
 FT /label="GCN4  
 FT /note="amino acids 249-281 of GCN4 LZ variant"  
 PN W09710843-A1.  
 PD 27-MAR-1997.  
 PE 20-SEP-1996; U5188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazoneis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer.  
 PS Disclosure: Refer to Page 8; 82pp; English.  
 CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
 CC of human wild-type p53 tumour suppressor (see also W13948) linked  
 CC to a C-terminal portion of the LZ variant (see also W13955) of  
 CC GCN4 and, in some cases, the C-terminal portion of wild-type  
 CC p53. The chimeric proteins have DNA binding activity and can  
 CC replace lost or insufficient p53 function, providing the means for  
 CC pharmacological rescue of p53 function in cancer patients. Nucleic  
 CC acids coding for modified p53 constructs can be used for cancer  
 CC gene therapy.  
 SO Sequence 361 AA;

Query Match 100.0%; Score 64; DB 21; Length 361;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 spqpkkkp1 323  
 Oy 1 SPOPKKKPL 9

RESULT 10  
 ID W28479 standard; Protein; 363 AA.  
 AC W28479;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-325 encoded by pEC114.  
 KM Leucine zipper domain; LZD; oligomerisation domain; mutant; mutin;  
 KM substitution; replacement; transactivation; viral protein VP16; HSV;  
 KM anti-oncogene; hyperproliferation; cancer; restenosis;  
 KM tumour suppression; apoptosis.  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 PN W09704092-A1.  
 PD 06-FEB-1997.  
 PE 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON) RHONE-POULENC RORER SA.  
 PI Bracco L, Concellier E;  
 DR WPI; 97-132633/12.  
 DR N-PSDB; T86215.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 30; Pages 76-78; 133p; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TAD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-325 and comprising  
 CC the VP16 TAD, amino acids 75-325 of human wild-type p53 and a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing

CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 24; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 323 spqpkkkp1 331  
 |||||  
 1 SPQPKKKPL 9

QY

RESULT 11  
 ID W28480 standard; Protein; 363 AA.  
 AC W28480;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutin;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Chimeric - Homo sapiens.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_difference 189  
 FT /note="Arg residue at position 182 of wild-type  
 FT p53 has been mutated to His"  
 PN MO9704092-A1.  
 PD 06-FEB-1997;  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON) RHONE-POULENC RORER, SA.  
 PI Bracco L, Icosseiller E;  
 DR WPT: 97-12263/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 30: Page -: 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-325H and comprising  
 CC the VP16 TD, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant V-325).  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 24; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 323 spqpkkkp1 331  
 |||||  
 1 SPQPKKKPL 9

QY

RESULT 12  
 ID W13971 standard; Protein; 363 AA.  
 AC W13971;

DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53R284del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN MO9710843-A1.  
 PD 27-MAR-1997;  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPT: 97-202618/18.

PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1: 51-52; 82pp; English.  
 CC Modified p53 variant p53R284del364-393 (W13971) has a Thr284 to Arg  
 CC substn. (see also W13949) and a deletion of the C-terminal 30  
 CC amino acids. The R284K substitution, introduced by site-directed  
 CC mutagenesis of p53 DNA, provides a novel p53-DNA contact between a  
 CC phosphate of the DNA backbone and p53. The C-terminal deletion  
 CC permits in vitro DNA binding. The variant provides the means for  
 CC pharmacological rescue of p53 function in cancer patients. Other  
 CC modified p53 constructs (W13949-50, W13953-54, W13968-77) have also  
 CC been produced. Nucleic acids coding for modified p53 can be used  
 CC for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 spqpkkkp1 323  
 |||||  
 1 SPQPKKKPL 9

QY

RESULT 13  
 ID W13974 standard; Protein; 363 AA.  
 AC W13974;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53H273del1364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN MO9710843-A1.  
 PD 27-MAR-1997;  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPT: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1: 56-57; 82pp; English.  
 CC Modified p53 variant p53H273del1364-393 (W13974) has the tumour-  
 CC derived histidine 273 mutation (see also W13952) and a deletion  
 CC of the C-terminal 30 amino acids of wild-type p53 (see also  
 CC W13948). His273 is a Class I p53 tumour mutation that affects DNA  
 CC binding. The C-terminal deletion, introduced by site-directed  
 CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
 CC tumour mutant. This provides the means for pharmacological rescue  
 CC of p53 function in cancer patients. Other modified p53 constructs  
 CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
 CC acids coding for modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 7.06e-01;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 spqpkkkp1 323

OY 1 SPOPKKKPL 9

RESULT 15

ID W13973 standard; Protein; 363 AA.  
AC W13973;  
DE 25-JUN-1997 (first entry)  
DE Modified p53 variant p53Q248R284del364-393.  
KM p53; tumour suppressor; cancer; therapy; cell proliferation;  
KM apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-A1.  
PD 27-MAR-1997.  
PE 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PI (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazoneis TD;  
DR WPI: 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in treatment of cancer.  
PS Example 1; 34-36; 82pp; English.  
CC Modified p53 variant p53Q248R284del364-393 (W13973) has the tumour-derived Gln248 mutation (see also W13951), a Thr284 to Arg substitution (see also W13949) and a deletion of the 30 C-terminal amino acids of wild-type p53 (W13948). Gln248 is a Class I p53 tumour mutation that affects DNA binding. The T284R substitution, introduced by site-directed mutagenesis of p53 DNA, provides a novel p53-DNA contact between a phosphate of the DNA backbone and p53, and restores DNA binding. The C-terminal deletion permits in vitro DNA binding. The construct provides the means for pharmacological rescue of p53 function in cancer patients. Other modified p53 constructs (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic acids coding for modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 7.06e-01; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 315 spqpkkkpl 323

OY 1 SPOPKKKPL 9

RESULT 15

ID W13972 standard; Protein; 363 AA.  
AC W13972;  
DE 25-JUN-1997 (first entry)  
DE Modified p53 variant p53Q248del364-393.  
KM p53; tumour suppressor; cancer; therapy; cell proliferation;  
KM apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-A1.  
PD 27-MAR-1997.  
PE 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PI (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazoneis TD;  
DR WPI: 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in treatment of cancer.  
PS Example 1; 53-54; 82pp; English.  
CC Modified p53 variant p53Q248del364-393 (W13972) has the tumour-derived glutamine 248 mutation (see also W13951) and a deletion of the C-terminal 30 amino acids of wild-type p53 (see also W13948). Gln248 is a Class I p53 tumour mutation that affects DNA binding. The C-terminal deletion, introduced by site-directed mutagenesis of p53 DNA, activates the DNA binding of the p53 tumour mutant. This provides the means for pharmacological rescue

CC of p53 function in cancer patients. Other modified p53 constructs (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic acids coding for modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 7.06e-01; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 315 spqpkkkpl 323

OY 1 SPOPKKKPL 9

Search completed: Sun Sep 13 10:41:52 1998  
Job time : 23 secs.



A25397; B25397; S42452; S42453; I38082; I38083; I38084;  
I38085; I38086; I38087; I38088; I38089; I38090; I38091;  
I38092; I38093; A44905; I58354; I78850; S60133

REFERENCE  
#authors Lamb, P.; Crawford, L.  
#journal Mol. Cell. Biol. (1986) 6:1379-1385  
#title Characterization of the human p53 gene.  
#cross-references MUID:87064416  
#accession A25224

##molecule-type DNA  
##residues 1-393 ##label LAM  
##cross-references EMBL:X01405; GB:M13121; GB:N00032; NID:9189460;  
PDB:9386994

REFERENCE  
#authors Buchman, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.;  
Georgiev, G.P.  
#journal Gene (1988) 70:245-252  
#title A variation in the structure of the protein-coding region of  
the human p53 gene.  
#cross-references MUID:89108008  
#accession A43073

##molecule-type DNA  
##residues 1-393 ##label BUC  
##note this 72-Arg allele appears to be about 5 times more  
frequent than the 72-Pro allele

#accession J10436

##molecule-type DNA  
##residues 1-71, 'P', '73-393' ##label BU2  
##cross-references EMBL:X2898; NID:9189474; PID:9189476  
##note this 72-Pro allele was found in both normal and  
malignant cell lines

REFERENCE  
#authors Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.  
#submission submitted to the EMBL Data Library, August 1990  
#accession S40773

##molecule-type DNA  
##residues 1-393 ##label CHU  
##cross-references EMBL:X54156; NID:935213; PID:935214

REFERENCE  
#authors Matlashevski, G.; Lamb, P.; Plm, D.; Peacock, J.; Crawford,  
L.; Benham, S.  
#journal EMBO J. (1984) 3:3257-3262  
#title Isolation and characterization of a human p53 cDNA clone:  
expression of the human p53 gene.  
#accession S42669

##molecule-type mRNA  
##status 101-393 ##label MK1  
##cross-references EMBL:X01405; NID:935215; PID:9642241

REFERENCE  
#authors Zakut-Houri, R.; Biernacki, B.; Givol, D.; Oren, M.  
#journal EMBO J. (1985) 4:1251-1255  
#title Human p53 cellular tumor antigen: cDNA sequence and  
expression in COS cells.  
#cross-references MUID:85230577  
#accession A22837

##molecule-type mRNA  
##residues 1-71, 'P', '73-393' ##label ZAK  
##cross-references EMBL:X02469; EMBL:M6050; NID:935209; PID:935210

REFERENCE  
#authors Harlow, E.; Williamson, N.M.; Raiston, R.; Helfman, D.M.;  
Adams, T.E.  
#journal Mol. Cell. Biol. (1985) 5:1601-1610  
#title Molecular cloning and in vitro expression of a cDNA clone for  
human cellular tumor antigen p53.  
#accession A55060

##molecule-type mRNA  
##residues 1-71, 'P', '73-272', 'H', '274-393' ##label HA3  
##cross-references GB:K03199; NID:9189478; PID:9189479  
##experimental-source clone pR4-2, cell line A431

REFERENCE  
#authors Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.;  
Arai, N.; Rotter, V.  
#journal Mol. Cell. Biol. (1986) 6:4650-4656

#title Molecular basis for heterogeneity of the human p53 protein.  
#cross-references MUID:87089826  
#accession A25397

##molecule-type mRNA  
##residues 1-78, 'T', '80-393' ##label HAR  
##cross-references EMBL:M14694; NID:933813; PID:933814  
##experimental-source clone p53-H-1, transformed hybridoma SV-80 cell  
line

#accession B25397

##molecule-type mRNA  
##residues 1-71, 'P', '73-78', 'T', '80-393' ##label HA2  
##cross-references EMBL:M14695; NID:933815; PID:933816  
##experimental-source clone p53-H-19, transformed hybridoma SV-80 cell  
line

REFERENCE  
#authors Matlashevski, G.J.; Tuck, S.; Plm, D.; Lamb, P.; Schneider,  
J.; Crawford, L.V.  
#journal Mol. Cell. Biol. (1987) 7:961-963  
#title Primary structure polymorphism at amino acid residue 72 of  
human p53.  
#accession S42452

##molecule-type mRNA: DNA  
##residues 66-71, 'P', '73-79' ##label MK2  
##experimental-source clone lambda C113  
##note 72-Cys was also found, and appears to represent a  
polymorphism

#accession S42453

##molecule-type mRNA: DNA  
##residues 66-79 ##label MAT  
##experimental-source clone J6K

REFERENCE  
#authors Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vonsden, K.H.;  
Crook, T.; J. (1991) 10:2879-2887  
#journal EMBO J. (1991) 10:2879-2887  
#title p53 is frequently mutated in Burkitt's lymphoma cell lines.  
#cross-references MUID:92007731  
#accession I38082

##status translated from GB/EMBL/DBD  
##molecule-type mRNA  
##residues 1-189, 'LSTLSSEKKEICVWSIMTETLEFDIWMCPMSRLRLAT',  
'VPPSTTTCVTPANAA' ##label F01  
##cross-references EMBL:X60010; NID:9506432; PID:9506433  
##note deletion of a C nucleotide causes a frameshift at  
position 566

#accession I38083

##status translated from GB/EMBL/DBD  
##molecule-type mRNA  
##residues 1-192, 'R', '194-393' ##label F02  
##cross-references EMBL:X60011; NID:9506434; PID:9506435

#accession I38084

##status translated from GB/EMBL/DBD  
##molecule-type mRNA  
##residues 1-393 ##label F03  
##cross-references EMBL:X60012; NID:9506436; PID:9506437

#accession I38085

##status translated from GB/EMBL/DBD  
##molecule-type mRNA  
##residues 1-245, 'T', '247-393' ##label F04  
##cross-references EMBL:X60013; NID:9506438; PID:9506439

#accession I38086

##status translated from GB/EMBL/DBD  
##molecule-type mRNA  
##residues 1-236, 'T', '238-393' ##label F05  
##cross-references EMBL:X60014; NID:9506440; PID:9506441

#accession I38087

##status translated from GB/EMBL/DBD  
##molecule-type mRNA  
##residues 1-247, 'Q', '249-393' ##label F06  
##cross-references EMBL:X60015; NID:9506442; PID:9506443

#accession I38088

##status translated from GB/EMBL/DBD  
##molecule-type mRNA  
##residues 1-71, 'P', '73-237', 'Y', '239-393' ##label F07

```

#molecule_type mRNA
##residues 1-393 ##label RIG
##cross-references EMBL:X16384; NID:g22795; PID:g22796
CLASSIFICATION
#superfamily cellular tumor antigen p53
KEYWORDS
apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc

FEATURE
176,179,238,242 #binding_site zinc (Cys, His, Cys, Cys) #status
392 #predicted\
#binding_site phosphoryl-RNA (Ser) (covalent) #status
SUMMARY
#length 393 #molecular_weight 43696 #checksum 4263

Query Match 100.0%; Score 64; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.92e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323
|||||||
QY 1 SPQPKKKPL 9

RESULT 4
ENTRY JC6193 #type complete
TITLE tumor suppressor p53 - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
08-Sep-1997
ACCESSIONS
REFERENCE
#author Le Goas, F.; May, P.; Ronco, P.; de Fromental, C.C.
#journal Gene (1997) 185:169-173
#title cDNA cloning and immunological characterization of rabbit
p53.
#accession JC6193
#molecule_type mRNA
##residues 1-391 ##label LEA
##cross-references EMBL:X90592; NID:g1532043; PID:el94962; PID:g1532044

GENETICS
#gene p53
CLASSIFICATION
#superfamily cellular tumor antigen p53
KEYWORDS
tumor
SUMMARY
#length 391 #molecular_weight 43435 #checksum 4367

Query Match 87.5%; Score 56; DB 2; Length 391;
Best Local Similarity 88.9%; Pred. No. 1.15e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 313 SPQTKKKPL 321
|||||||
QY 1 SPQTKKKPL 9

RESULT 5
ENTRY S02192 #type complete
TITLE cellular tumor antigen p53 - rat
ALTERNATE_NAMES
gene p53 protein; nuclear oncoprotein p53
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change
08-Sep-1997
ACCESSIONS
REFERENCE
#author Soussi, T.; de Fromental, C.C.; Breugnot, C.; May, E.
#journal Nucleic Acids Res. (1988) 16:11384
#title Nucleotide sequence of a cDNA encoding the rat p53 nuclear
oncoprotein.
#cross-references MIMD:89083585
#accession S02192
#molecule_type mRNA
##residues 1-391 ##label SOU
##cross-references EMBL:X13058; NID:g56828; PID:g56829

```

```

REFERENCE      S41149
#authors      Hulla, J.E.; Schneider, R.P.
#journal      Nucleic Acids Res. (1993) 21:713-717
#title        Structure of the rat p53 tumor suppressor gene.
#accession    S41149
#status       preliminary; nucleic acid sequence not shown;
               translation not shown
#molecule_type DNA
##residues    1-173, 'W', 175-391 ##label HUTL
##cross-references EMBL:107909
##note         the nucleotide sequence was submitted to the EMBL Data
               Library, December 1992

GENETICS
#introns      25/2: 32/3; 123/3; 185/1; 259/2; 305/1; 329/3; 365/2
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS       apoptosis; cell division control; DNA binding; homotetramer;
               nucleus; phosphoprotein; transcription regulation; tumor
               suppressor; zinc

FEATURE
174,177,236,240 #binding_site zinc (Cys, His, Cys, Cys) #status
                 predicted\
390              #binding_site phosphoryl-RNA (Ser) (covalent) #status
                 predicted\
SUMMARY         #length 391 #molecular-weight 43451 #checksum 7105
Query Match    87.5%; Score 56; DB 2; Length 391;
Best Local Similarity 88.9%; Pred. No. 1.15e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 313 SPOPKRKL 321
11111111
OY 1 SPOPKRKL 9

RESULT 6
ENTRY      JH0633 #type complete
#authors    Legros, Y.; McIntyre, P.; Soussi, T.
#journal     Gene (1992) 112:247-250
#title       The cDNA cloning and immunological characterization of
               hamster p53.
#accessions JH0633
REFERENCE    JH0633
#journal     Legros, Y.; McIntyre, P.; Soussi, T.
#title       Gene (1992) 112:247-250
#accession   JH0633
#molecule_type mRNA
##residues   1-396 ##label LEG
##cross-references GB:M75144; NID:g191414; PID:g191415
#note        ##experimental_source kidney, strain MPl

GENETICS
#gene        p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS       apoptosis; cell division control; DNA binding; homotetramer;
               nucleus; phosphoprotein; transcription regulation; tumor
               suppressor; zinc

FEATURE
179,182,241,245 #binding_site zinc (Cys, His, Cys, Cys) #status
                 predicted\
395              #binding_site phosphoryl-RNA (Ser) (covalent) #status
                 predicted\
SUMMARY         #length 396 #molecular-weight 43631 #checksum 6617
Query Match    82.8%; Score 53; DB 2; Length 396;
Best Local Similarity 77.8%; Pred. No. 5.02e-01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 318 SPOPKRKL 326
11111111
OY 1 SPOPKRKL 9

```

```

RESULT 7
ENTRY      S75263 #type complete
#authors    hypothetical protein - Synecocystis sp. (PCC 6803)
#journal     #formal_name Synecocystis sp.
#accession  PCC 6803
#status     25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
               09-Sep-1997
#molecule_type DNA
##residues   S75263
##cross-references EMBL:D90904; NID:g1652225; PID:d1017910; PID:g1652254
#note        the nucleotide sequence was submitted to the EMBL Data
               Library, June 1996

GENETICS
#start_codon GTG
SUMMARY      #length 209 #molecular-weight 23287 #checksum 3304
Query Match  76.6%; Score 49; DB 2; Length 209;
Best Local Similarity 44.4%; Pred. No. 3.33e-00;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 122 PPHRRRPL 130
11111111
OY 1 SPOPKRKL 9

RESULT 8
ENTRY      S38824 #type complete
#authors    Aral, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
               Shohat, O.; Rotter, V.
#journal     Mol. Cell. Biol. (1986) 6:3232-3239
#title       Immunologically distinct p53 molecules generated by
               alternative splicing.
#accessions S38824
REFERENCE    S38824
#journal     Aral, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
               Shohat, O.; Rotter, V.
#title       Mol. Cell. Biol. (1986) 6:3232-3239
#accession   S38824
#molecule_type mRNA
##residues   S35478
##cross-references EMBL:M13874; NID:g200202; PID:g200203
#note        the nucleotide sequence was submitted to the EMBL Data
               Library, July 1988

COMMENT      This sequence, produced by alternative splicing of the tenth
               intron, lacks the carboxyl-terminal sequence necessary for

```



covalent attachment of RNA. The function of this minor splice form is not known.

CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS alternative splicing; phosphoprotein

## FEATURE

1-44 #domain transcription activation #status predicted  
#label TRA  
16-26 #region conserved region I  
99-289 #domain DNA-binding core #status predicted #label DBC  
108-121 #region L1 loop  
114-139 #region L2 loop  
160-192 #region conserved region II  
168-178 #region conserved region III  
231-252 #region conserved region IV  
233-248 #region L3 loop  
267-283 #region conserved region V  
313-319 #region nuclear location signal  
319-357 #region tetramer association  
7,9,12,18,23,37 #binding\_site phosphate (Ser) (covalent) #status predicted  
173,176,235,239 #binding\_site zinc (Cys, His, Cys, Cys) #status predicted  
312 #binding\_site phosphate (Ser) (covalent) (by cdc2 kinase) #status predicted

SUMMARY #length 381 #molecular-weight 42498 #checksum 8703

Query Match 76.6% Score 49; DB 2; Length 381;  
Best Local Similarity 77.8%; Pred. No. 3.33e+00;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 312 SPPQKKPL 320

OY 1 SPPQKKPL 9

## RESULT 9

ENTRY #type complete  
TITLE cellular tumor antigen p53 - mouse  
ALTERNATE\_NAMES oncoprotein p53  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 28-Aug-1985 #sequence\_revision 04-Oct-1996 #text\_change 05-Sep-1997

ACCESSIONS A22739; S06336; A02684; S38822; S38823; I48703

## REFERENCE

#authors A22739  
#journal Biernz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.  
#cross-references M01D:8502173  
#accession A22739

#molecule\_type DNA  
#residues 1-134, 'V', 136-390 #label B1E

## REFERENCE

#authors S06336  
#journal Chumakov, P.M.  
#cross-references M01D:88221682  
#accession S06336

#status not compared with conceptual translation

#molecule\_type mRNA  
#residues 1-134, 'V', 136-390 #label CHU  
#accession A02684

## REFERENCE

#authors Zakut-Houri, R.; Oren, M.; Biernz, B.; Lavie, V.; Hazum, S.; Givol, D.  
#journal Nature (1983) 306:594-597  
#title A single gene and a pseudogene for the cellular tumour antigen p53.  
#cross-references M01D:84068204  
#accession A02684

#molecule\_type mRNA

#residues 1-159, 'H', 161-167, 'G', 169-233, 'I', 235-390 #label ZAK

## REFERENCE

#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.; Shohat, O.; Rotter, V.

#journal Mol. Cell. Biol. (1986) 6:3232-3239  
#title Immunologically distinct p53 molecules generated by alternative splicing.

#accession S38822

#status preliminary

#molecule\_type mRNA

#residues 1-390 #label ARA

#cross-references EMBL:M13872; NID:g200198; PID:g200199

#accession S38823

#status preliminary

#molecule\_type mRNA

#residues 1-167, 'G', 169-233, 'I', 235-390 #label AR2

#cross-references EMBL:M13873

#accession I48703

#status preliminary

#molecule\_type mRNA

#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES

#cross-references EMBL:X00741; NID:g53570; PID:g53571

#accession I48703

#status preliminary

#molecule\_type mRNA

#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES

#cross-references EMBL:X00741; NID:g53570; PID:g53571

#accession I48703

#status preliminary

#molecule\_type mRNA

#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES

#cross-references EMBL:X00741; NID:g53570; PID:g53571

#accession I48703

#status preliminary

#molecule\_type mRNA

#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES

#cross-references EMBL:X00741; NID:g53570; PID:g53571

#accession I48703

#status preliminary

#molecule\_type mRNA

#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES

#cross-references EMBL:X00741; NID:g53570; PID:g53571

#accession I48703

#status preliminary

#molecule\_type mRNA

#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES

#cross-references EMBL:X00741; NID:g53570; PID:g53571

#accession I48703

#status preliminary

#molecule\_type mRNA

#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES

#cross-references EMBL:X00741; NID:g53570; PID:g53571

#accession I48703

#status preliminary

#molecule\_type mRNA

#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES

#cross-references EMBL:X00741; NID:g53570; PID:g53571

#accession I48703

#status preliminary

#molecule\_type mRNA

#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES

#cross-references EMBL:X00741; NID:g53570; PID:g53571

#accession I48703

#status preliminary

#molecule\_type mRNA

#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES

#cross-references EMBL:X00741; NID:g53570; PID:g53571

#accession I48703

#status preliminary

#molecule\_type mRNA

#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES

#cross-references EMBL:X00741; NID:g53570; PID:g53571

ACCESSIONS JC6176  
 REFERENCE JC6176  
 #authors Lee, H.; Larner, J.M.; Hamlin, J.L.  
 #journal Gene (1997) 184:177-183  
 #title Cloning and characterization of Chinese hamster p53 cDNA.  
 #contents liver  
 #accession JC6176  
 ##molecule\_type mRNA  
 ##residues 1-393 ##label LEE  
 ##cross-references GB:050395; NID:91842229; PID:91842230  
 COMMENT This protein is a multimer, it plays the central role in a complex DNA damage-sensing network. It binds to replication factor and DNA-binding protein, and affects DNA replication, transcription, and recombination by protein/protein interactions.

GENETICS  
 #gene p53  
 CLASSIFICATION #superfamily cellular tumor antigen p53  
 KEYWORDS liver; tumor  
 SUMMARY #length 393 #molecular-weight 43362 #checksum 4043

Query Match 76.6%; Score 49; DB 2; Length 393;  
 Best Local Similarity 77.8%; Pred. No. 3.32e+00;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 315 SPPKKTL 323  
 1111111  
 Oy 1 SPOPKKPL 9

RESULT 11  
 ENTRY A57416 #type complete  
 TITLE ribosomal protein L7e, cytosolic - fruit fly (Drosophila melanogaster)  
 ORGANISM #formal\_name Drosophila melanogaster  
 DATE 08-Feb-1996 #sequence\_revision 08-Feb-1996 #text\_change 14-Nov-1997

C. ACCESSIONS A57416  
 #REFERENCE A57416  
 #authors Armes, N.; Fried, M.  
 #journal Mol. Cell. Biol. (1995) 15:2367-2373  
 #title The genomic organization of the region containing the Drosophila melanogaster rpl7a (Surf-3) gene differs from those of the mammalian and avian surfelt loci.  
 #accession A57416  
 ##status preliminary  
 ##molecule\_type DNA  
 ##residues 1-272 ##label ARM  
 ##cross-references GB:X82782  
 #note authors translated the codon GCG for residue 112 as Pro, AAC for residue 116 as Lys, and GNG for residue 117 as Leu

GENETICS  
 #gene rpl7a  
 ##cross-references FlyBase:FBgn0014026  
 #introns 5/3: 47/1; 170/3  
 CLASSIFICATION #superfamily rat ribosomal protein L7a  
 KEYWORDS protein biosynthesis; ribosome  
 SUMMARY #length 272 #molecular-weight 30732 #checksum 517

Query Match 73.4%; Score 47; DB 2; Length 272;  
 Best Local Similarity 75.0%; Pred. No. 8.32e+00;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 6 PRPKKPV 13  
 1111111  
 Oy 2 POPKKPL 9

RESULT 12  
 ENTRY S14915 #type complete  
 TITLE hypothetical protein 1 - Trypanosoma brucei gambiense  
 ORGANISM #formal\_name Trypanosoma brucei gambiense

DATE 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 08-Sep-1997  
 ACCESSIONS S14915; S08043  
 REFERENCE S14915  
 #authors Aksy, S.; Williams, S.; Chang, S.; Richards, F.F.  
 #journal Nucleic Acids Res. (1990) 18:785-792  
 #title SLACS retrotransposon from Trypanosoma brucei gambiense is similar to mammalian LINES.  
 ##cross-references MUID:90192150  
 #accession S14915  
 ##molecule\_type DNA  
 ##residues 1-404 ##label AKS  
 ##cross-references EMBL:X17078; NID:910533; PID:910534  
 KEYWORDS DNA binding; zinc finger  
 SUMMARY #length 404 #molecular-weight 45463 #checksum 6169

Query Match 73.4%; Score 47; DB 2; Length 404;  
 Best Local Similarity 75.0%; Pred. No. 8.32e+00;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 167 TPQKKKA 174  
 1111111  
 Oy 1 SPOPKKP 8

RESULT 13  
 ENTRY S2850A #type complete  
 TITLE stage 0 sporulation protein spo0A - Bacillus subtilis  
 ALTERNATE\_NAMES sporulation initiation two-component response regulator spo0A  
 ORGANISM #formal\_name Bacillus subtilis  
 DATE 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 20-Mar-1998

C. ACCESSIONS A94036; A22665; 140013; A69710; A26068; A29099; B22665  
 #REFERENCE A94036  
 #authors Ferrari, F.A.; Trach, K.; Lecocq, D.; Spence, J.; Ferrari, E.; Hoch, J.A.  
 #journal Proc. Natl. Acad. Sci. U.S.A. (1985) 82:2647-2651  
 #title Characterization of the spo0A locus and its deduced product.  
 ##cross-references MUID:85190553  
 #accession A94036  
 ##molecule\_type DNA  
 ##residues 1-267 ##label FER  
 ##cross-references GB:M10082; NID:9143584; PID:9143585

REFERENCE A22665  
 #authors Kudoh, J.; Ikeuchi, T.; Kurahashi, K.  
 #journal Proc. Natl. Acad. Sci. U.S.A. (1985) 82:2665-2668  
 #title Nucleotide sequences of the sporulation gene spo0A and its mutant genes of Bacillus subtilis.  
 ##cross-references MUID:85190557  
 #accession A22665  
 ##molecule\_type DNA  
 ##residues 1-267 ##label KUD  
 ##cross-references GB:M10082; NID:9143584; PID:9143585  
 #note these authors assume that the codon ATG for Met-29 is the initiator for translation

REFERENCE A26068  
 #authors Ikeuchi, T.; Kudoh, J.; Tsunasawa, S.  
 #journal Mol. Gen. Genet. (1986) 203:371-376  
 #title Amino-terminal structure of spo0A protein and sequence homology with spoof and spoOB proteins.  
 ##cross-references MUID:86310272  
 #note annotation  
 #contents Initiation at the codon GTG for Met-1 was demonstrated  
 #note I40013  
 #authors Shoji, K.; Hiratsuka, S.; Kawamura, F.; Kobayashi, Y.  
 #journal J. Gen. Microbiol. (1988) 134:3249-3257  
 #title New Suppressor Mutation surOB of spoOB and spoOF Mutations in Bacillus subtilis.  
 ##cross-references MUID:90063528  
 #accession I40013  
 ##status translated from GB/EMBL/DBJ  
 ##molecule\_type DNA  
 ##residues 1-16 ##label RES

##cross-references GB:M23656; NID:g143720; PID:g143721  
REFERENCE A69580

## #authors

Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, B.; Bourcier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Dusterhoft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Haelech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kashara, Y.; Klier-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Paro, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Presecan, E.; Puig, P.; Purnelle, B.; Rapoport, G.; Rev, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeder, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serron, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Taconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpe, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wandut, R.; Wedler, E.; Wedler, H.; Wetzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zaststein, E.; Yoshikawa, H.; Zanchin, A.

## #journal

The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

## #accession

A69710

## #status

nucleic acid sequence not shown; translation not shown

## #molecule\_type

DNA

## #residues

1-267 #label KUN

## #experimental\_source

strain 168

## #comment

This protein is involved in the initiation of sporulation.

## #genetics

sp00A

## #gene

map-position 215 (degrees)

## #start\_codon

GTG

## #classification

superfamily stage 0 sporulation protein A; response phosphoprotein; sporulation

## #keywords

regulator homology

## #feature

6-119

## #domain

response regulator homology #label RRH

## #binding\_site

phosphate (Aap) (covalent) #status

## #predicted

predicted

## #length

267 #molecular-weight 29691 #checksum 9619

## #query\_match

71.9%; Score 46; DB 1; Length 267;

## #best\_local\_similarity

75.0%; Pred. No. 1.30e+01;

## #matches

6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

## #db

146 PEPKKNL 153

## #oy

2 POPKKRPL 9

## #result

14

## #entry

S14959

## #title

Proline-rich protein - wheat

## #organism

#formal\_name Triticum aestivum

## #date

20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change

09-Sep-1997  
S14959

## #reference

S14959

## #authors

Raines, C.A.; Lloyd, J.C.; Chao, S.; John, U.P.; Murphy, G.J.P.

## #journal

Plant Mol. Biol. (1991) 16:663-670

## #title

A novel proline-rich protein from wheat.

## #cross-references

MUID:91329699

## #accession

S14959

## #status

preliminary

## #molecule\_type

mRNA

## #residues

1-378 #label RAI

## #cross-references

EMBL:X52472; NID:g21841; PID:g21842

## #length

378 #molecular-weight 42119 #checksum 8388

## #query\_match

71.9%; Score 46; DB 2; Length 378;

## #best\_local\_similarity

75.0%; Pred. No. 1.30e+01;

## #matches

6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

## #db

335 PEPKKNL 342

## #oy

2 POPKKRPL 9

## #result

15

## #entry

S74291

## #title

hypothetical protein YCR030C - yeast (Saccharomyces cerevisiae)

## #alternate\_names

hypothetical protein YCR029C

## #organism

formal name Saccharomyces cerevisiae

## #date

19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change

## #accessions

S74291; S40970; S19442; S19440

## #authors

Medler, H.; Wandut, R.

## #submitted

submitted to the Protein Sequence Database, September 1996

## #accession

S74291

## #molecule\_type

DNA

## #residues

1-870 #label WED

## #cross-references

EMBL:X59720; NID:g1907116; PID:e308993; PID:g1907173;

## #note

this is a revision to the sequence from reference S19439

## #reference

S25336

## #authors

Carbone, M.L.A.; Panzeri, L.; Falconi, M.M.; Carcano, C.; Plevani, P.; Lucchini, G.

## #journal

Yeast (1992) 8:805-812

## #title

Nucleotide sequence of 9.2 kb left of CRY1 on yeast chromosome III from strain AB972: evidence for a Ty insertion and functional analysis of open reading frame YCR28.

## #accession

S40970

## #status

translation not shown

## #molecule\_type

DNA

## #residues

203-719 'T', 721-823 'V', 825-830 'A', 832-870 #label CAR

## #cross-references

EMBL:S47818

## #reference

S19439

## #authors

Cederberg, H.; Hohmann, S.; Schaaff-Gerstenschlager, I.; Huse, K.; Zimmermann, F.K.

## #submitted

submitted to the Protein Sequence Database, March 1992

## #accession

S19442

## #molecule\_type

DNA

## #residues

1-652, 'LSVLI' #label CED

## #cross-references

EMBL:X59720; MIPS:YCR030C

## #note

this was assumed to be the complete sequence of protein YCR030C

## #accession

S19440

## #molecule\_type

DNA

## #residues

731-870 #label CEW

## #cross-references

EMBL:X59720

## #note

this sequence has been revised in reference S74291

## #genetics

YCR029C

## #title

this sequence has been revised in reference S74291

## #organism

YCR029C

## #date

20-Feb-1995

## #sequence

revision

## #text\_change

20-Feb-1995

## #sequence

revision

## #text\_change

20-Feb-1995

## #sequence

revision

## #text\_change

20-Feb-1995

Sun Sep 13 10:56:13 1998

US-08-452-843-19.rpr

Page 8

```
#map_position 3R
#note YCR030C
SUMMARY #length 870 #molecular-weight 96124 #checksum 3413

Query Match 71.9%; Score 46; DB 2; Length 870;
Best Local Similarity 75.0%; Pred. No. 1.30e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 483 POSKTKPL 490
11:1111
QY 2 POPKKKPL 9
11:1111

Search completed: Fri Sep 11 13:48:56 1998
Job time : 26 secs.
```

# MORSE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch.p protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Sep 11 13:49:13 1998; MasPar time 2.54 Seconds  
Tabular output not generated. 88.717 Million cell updates/sec

Title: >US-08-452-843-19  
Description: (1-9) from US08452843.pep  
Perfect Score: 64  
Sequence: 1 SPQPKKRL 9

Scoring table: PAM 150  
Gap 15

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swiss1

Statistics: Mean 21.734; Variance 23.588; scale 0.921

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID         | Description            | Pred. No. |
|------------|-------|-------------|--------|----|------------|------------------------|-----------|
| 1          | 64    | 100.0       | 314    | 1  | P53_SPEBE  | CELLULAR TUMOR ANTIGEN | 9.18e-05  |
| 2          | 64    | 100.0       | 386    | 1  | P53_BOVIN  | CELLULAR TUMOR ANTIGEN | 9.18e-05  |
| 3          | 64    | 100.0       | 393    | 1  | P53_HUMAN  | CELLULAR TUMOR ANTIGEN | 9.18e-05  |
| 4          | 64    | 100.0       | 393    | 1  | P53_CERAE  | CELLULAR TUMOR ANTIGEN | 9.18e-05  |
| 5          | 56    | 87.5        | 382    | 1  | P53_SHEEP  | CELLULAR TUMOR ANTIGEN | 1.23e-02  |
| 6          | 56    | 87.5        | 391    | 1  | P53_RABIT  | CELLULAR TUMOR ANTIGEN | 1.23e-02  |
| 7          | 56    | 87.5        | 391    | 1  | P53_RAT    | CELLULAR TUMOR ANTIGEN | 1.23e-02  |
| 8          | 53    | 82.8        | 396    | 1  | P53_MESU   | CELLULAR TUMOR ANTIGEN | 7.05e-02  |
| 9          | 49    | 76.6        | 280    | 1  | P53_HORSE  | CELLULAR TUMOR ANTIGEN | 6.63e-01  |
| 10         | 49    | 76.6        | 390    | 1  | P53_MOUSE  | CELLULAR TUMOR ANTIGEN | 6.63e-01  |
| 11         | 49    | 76.6        | 393    | 1  | P53_CRIGR  | CELLULAR TUMOR ANTIGEN | 6.63e-01  |
| 12         | 47    | 73.4        | 271    | 1  | PL7A_DROME | 60S RIBOSOMAL PROTEIN  | 1.95e+00  |
| 13         | 47    | 73.4        | 386    | 1  | P53_FELCA  | CELLULAR TUMOR ANTIGEN | 1.95e+00  |
| 14         | 47    | 73.4        | 404    | 1  | RTPI_TRYBG | RETROTRANSPOSABLE ELEM | 1.95e+00  |
| 15         | 46    | 71.9        | 50     | 1  | FLIZ_SALTY | FLIZ PROTEIN (FRAGMENT | 3.30e+00  |
| 16         | 46    | 71.9        | 267    | 1  | SPQA_BACSU | STAGE 0 SPOULATION PR  | 3.30e+00  |
| 17         | 46    | 71.9        | 657    | 1  | YCSO_YEAST | HYPOHETICAL 72.7 KD P  | 3.30e+00  |
| 18         | 46    | 71.9        | 1206   | 1  | FORM_MOUSE | FORMIN 4 (LIMB DEFORMI | 3.30e+00  |
| 19         | 46    | 71.9        | 1468   | 1  | FORM_MOUSE | FORMIN (LIMB DEFORMI   | 3.30e+00  |
| 20         | 46    | 71.9        | 2663   | 1  | CENR_HUMAN | CENTROMERIC PROTEIN E  | 3.30e+00  |
| 21         | 45    | 70.3        | 196    | 1  | RAC9_GOSHI | RAC-LIKE GTP BINDING P | 5.54e+00  |
| 22         | 45    | 70.3        | 747    | 1  | YHMA_CAEEL | HYPOHETICAL 83.2 KD P  | 5.54e+00  |
| 23         | 45    | 70.3        | 754    | 1  | ACSA_ACEXY | CELLULOSE SYNTHASE CAT | 5.54e+00  |

|    |    |      |      |   |            |                         |          |
|----|----|------|------|---|------------|-------------------------|----------|
| 24 | 45 | 70.3 | 1220 | 1 | DPOL_HAVER | DNA POLYMERASE (EC 2.7  | 5.54e+00 |
| 25 | 45 | 70.3 | 1259 | 1 | YTFN_ECOLI | HYPOHETICAL 136.8 KD    | 5.54e+00 |
| 26 | 44 | 68.8 | 71   | 1 | NXL1_NAJME | LONG NEUTROXIN 1 (NEU   | 9.24e+00 |
| 27 | 44 | 68.8 | 191  | 1 | SPQA_BACPU | STAGE 0 SPOULATION PR   | 9.24e+00 |
| 28 | 44 | 68.8 | 238  | 1 | NBMA_MAIZE | DNA-BINDING PROTEIN MN  | 9.24e+00 |
| 29 | 44 | 68.8 | 307  | 1 | YIDL_ECOLI | HYPOHETICAL TRANSCRIP   | 9.24e+00 |
| 30 | 44 | 68.8 | 451  | 1 | GAA2_HUMAN | GAMMA-AMINOBUTYRIC-ACI  | 9.24e+00 |
| 31 | 44 | 68.8 | 451  | 1 | GAA2_BOVIN | GAMMA-AMINOBUTYRIC-ACI  | 9.24e+00 |
| 32 | 44 | 68.8 | 451  | 1 | GAA2_MOUSE | GAMMA-AMINOBUTYRIC-ACI  | 9.24e+00 |
| 33 | 44 | 68.8 | 451  | 1 | GAA2_RAT   | GAMMA-AMINOBUTYRIC-ACI  | 9.24e+00 |
| 34 | 44 | 68.8 | 831  | 1 | SAS3_YEAST | PSM1 PROTEIN HOMOLOG 2  | 9.24e+00 |
| 35 | 44 | 68.8 | 862  | 1 | PMS2_HUMAN | PROTEIN-TYROSINE PHOSP  | 9.24e+00 |
| 36 | 44 | 68.8 | 1189 | 1 | PTNE_MOUSE | POSSIBLE GANGLIOCYTIC K | 1.53e+01 |
| 37 | 44 | 67.2 | 431  | 1 | GCVA_HYSEA | PROBABLE SENSOR PROTEI  | 1.53e+01 |
| 38 | 43 | 67.2 | 449  | 1 | YGYI_ECOLI | ZYXIN.                  | 1.53e+01 |
| 39 | 43 | 67.2 | 542  | 1 | ZYX_CHICK  | ZYXIN.                  | 1.53e+01 |
| 40 | 43 | 67.2 | 564  | 1 | ZYX_MOUSE  | ZYXIN.                  | 1.53e+01 |
| 41 | 43 | 67.2 | 587  | 1 | GGT_BACSU  | GAMMA-GLUTAMYLTRANSPEP  | 1.53e+01 |
| 42 | 43 | 67.2 | 1172 | 1 | PHYE_ARATH | PHYTOCHROME B.          | 1.53e+01 |
| 43 | 43 | 67.2 | 1967 | 1 | YGSO_YEAST | POTATIVE RNA HELICASE   | 1.53e+01 |
| 44 | 43 | 67.2 | 3110 | 1 | HD_RAT     | HUNTINGTIN (HUNTINGTON  | 1.53e+01 |
| 45 | 43 | 67.2 | 3119 | 1 | HD_MOUSE   | HUNTINGTIN (HUNTINGTON  | 1.53e+01 |

## ALIGNMENTS

RESULT 1  
ID P53\_SPEBE STANDARD; PRT; 314 AA.  
AC 064662;

DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).

GN TP53.  
OS SPERMOPHILUS BEECHERI (BEECHER GROUND SQUIRREL).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.

RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=THYMUS;  
RX MEDLINE: 95007566.

RA RIVKINA M.B., CULLEN J.M., ROBINSON W.S., MARION P.L.;  
CANCER RES. 54:5430-5437(1994).

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.

CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC DR EMBL: U43902; G1165312; -  
CC PROSITE: PS00348; P53; 1.  
CC ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
CC NUCLEAR PROTEIN; PHOSPHORYLATION.

KW NON\_TER  
FT DOMAIN 289 301 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT NON\_TER 314 314  
SQ SEQUENCE 314 AA; 34618 MW; D07F433B CRC32;

Query Match 100.0%; Score 64; DB 1; Length 314;  
Best Local Similarity 100.0%; Pred. No. 9.18e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 293 SPQPKKRL 301  
|||||||

OY 1 SPOPKKKPL 9

RESULT 2

ID P53-BOVIN STANDARD: PRT: 386 AA.

AC Q29628;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53.

GN TP53

OS BOS TAURUS (BOVINE);

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; ARTIODACTYLA.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-LIVER;

RX MEDLINE; 95352829.

RA DEQUIEDT F., KETTMANN R., BURRY A., WILLEMS L.;

RL DNA SEQ. 5:261-264(1995).

RN [2]

RP SEQUENCE OF 13-386 FROM N.A.

RC STRAIN-HOLSTEIN; TISSUE-THYMUS;

RX MEDLINE; 96401400.

RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;

RL VET. IMMUNOL. IMMUNOPATHOL. 52:53-63(1996).

RN [1]

RP FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.

CC [1]

CC SUBCELLULAR LOCATION: NUCLEAR.

CC [1]

CC DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.

CC [1]

CC SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

DR EMBL; X81704; G602333;

DR EMBL; D49825; G1729419;

DR PROSITE; PS00348; P53: 1.

KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.

FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).

FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

FT MOD.RES 385 385 PHOSPHORYLATION (BY SIMILARITY).

FT CONFLICT 380 380 R->T (IN REF. 2).

SO SEQUENCE 386 AA; 43255 MW; 0322B53D CRC32;

Query Match 100.0%; Score 64; DB 1; Length 386;

Best Local Similarity 100.0%; Pred. No. 9.18e-05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 308 SPOPKKKPL 316

OY 1 SPOPKKKPL 9

RESULT 3

ID P53-HUMAN STANDARD: PRT: 393 AA.

AC P04637;

DT 13-AUG-1987 (REL. 05, CREATED)

DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).

GN TP53

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 85230577.

RA ZAKUT-HOURI R., BIENZ-TADMOR B., GIVOL D., OREN M.;

RL EMO J. 4:1251-1255(1985).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 87064416.

RA LAMB P., CRAWFORD L.;

RL MOL. CELL. BIOL. 6:1379-1385(1986).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE; 85267676.

RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;

RL MOL. CELL. BIOL. 5:1601-1610(1985).

RN [4]

RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.

RX MEDLINE; 87089826.

RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N., ROTTNER V.;

RL MOL. CELL. BIOL. 6:4650-4656(1986).

RN [5]

RP SEQUENCE FROM N.A.

RX MEDLINE; 89108008.

RA BUCHMAN V.L., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P., GEORGIEV G.P.;

RL GENE 70:245-252(1988).

RN [6]

RP SEQUENCE OF 101-393 FROM N.A.

RX MEDLINE; 85126934.

RA MATIASHEWSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L., BECHTELMO S.;

RL EMO J. 3:3257-3262(1984).

RN [7]

RP NUCLEAR LOCALIZATION SIGNAL.

RX MEDLINE; 90191730.

RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;

RL ONCOGENE 5:423-426(1990).

RN [8]

RP STRUCTURE BY NMR OF 319-360.

RX MEDLINE; 94294808.

RA CLORE G.M., OMICHINSKI J.G., SARAGUCHI K., ZAMBRANO N., SAKAMOTO H., APPELBA E., GROENENBORN A.M.;

RL SCIENCE 265:386-391(1994).

RN [9]

RP STRUCTURE BY NMR OF 325-355.

RX MEDLINE; 95292092.

RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;

RL NAT. STRUCT. BIOL. 1:877-890(1994).

RN [10]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.

RX MEDLINE; 94294806.

RA CHO Y., GORINA S., JEFFERY P.D., PAVLETICH N.P.;

RL SCIENCE 265:346-355(1994).

RN [11]

RP REVIEW.

RX MEDLINE; 94090335.

RA HARRIS C.C.;

RL SCIENCE 262:1980-1981(1993).

RN [12]

RP REVIEW ON VARIANTS.

RX MEDLINE; 91289156.

RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;

RL SCIENCE 253:49-53(1991).

RN [13]

RP REVIEW ON VARIANTS.

RX MEDLINE; 96271983.

RA DE VRIES E.M.G., RICKE D.O., DE VRIES T.N., HARTMANN A., BLASSYK H., LIRO D., SOUSST T., KOVACH J.S., SOMMER S.S.;

RL HUM. MUTAT. 7:202-213(1996).

RN [14]

RP VARIANT ARG-72.

RX MEDLINE; 91153807.

RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;

Sun Sep 13 10:56:14 1998

RL HUM. GENET. 86:369-370(1991).  
 RN [15]  
 RP VARIANT LFS THR-133.  
 RX MEDLINE: 92034774.  
 RA LAM J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
 RL CANCER RES. 51:6385-6387(1991).  
 RN [16]  
 RP VARIANT LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
 RX MEDLINE: 91057657.  
 RA MALIKIN D., LI F.P., STRONG L.C., FRAMMENTI J.F. JR., NELSON C.E.,  
 RA KIM D.H., KASSEL J., GRITKA M.A., BISCHOFF F.Z., TAINISKY M.A.,  
 RA FRIEND S.H.;  
 RL SCIENCE 250:1233-1238(1990).  
 RN [17]  
 RP VARIANT LFS ASP-245.  
 RX MEDLINE: 91080929.  
 RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
 RL NATURE 348:747-749(1990).  
 RN [18]  
 RP VARIANT LFS LEU-272.  
 RX MEDLINE: 92147883.  
 RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
 RA POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
 RA KNOTSEN T., MINNA J.D.;  
 RL J. CLIN. INVEST. 89:640-647(1992).  
 RN [19]  
 RP VARIANT LFS HIS-273 AND VAL-325.  
 RX MEDLINE: 92228023.  
 RA MALIKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
 RA GGBHARDT M.C., ANDERSEN T.I., BORRESSEN A.-L., LI F.P., GARBER J.,  
 RA STRONG L.C.;  
 RL NEW ENGL. J. MED. 326:1309-1315(1992).  
 RN [20]  
 RP VARIANT BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
 RX MEDLINE: 90295284.  
 RA BARBER J., IGGO R., GANNON J., LANE D.P.;  
 RL ONCOGENE 5:893-899(1990).  
 RN [21]  
 RP VARIANT COLON TUMORS PHE-241 AND HIS-273.  
 RX MEDLINE: 91017544.  
 RA RODRIGUES N.R., ROMAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
 RA GANNON J.V., LANE D.P.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 87:7555-7559(1990).  
 RN [22]  
 RP VARIANT COLORECTAL CANCER MUTATIONS.  
 RX MEDLINE: 91282784.  
 RA ISHIKAWA C., SATO T., GAMOH M., SUZUKI T., SHIBATA H., KANMARU R.,  
 RA WAKU A., YAMAZAKI T.;  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 177:901-906(1991).  
 RN [23]  
 RP VARIANT OSOEPHAGUS TUMORS L-155; A-155; H-175; F-176 AND H-273.  
 RX MEDLINE: 91330175.  
 RA CASSON A.G., MUKHOPADHYAY T., CLEARY K.R., RO J.Y., LEVIN B.,  
 RA ROTH J.A.;  
 RL CANCER RES. 51:4495-4499(1991).  
 RN [24]  
 RP VARIANT HEPATOCELLULAR CARCINOMAS MUTATIONS IN CHINA.  
 RX MEDLINE: 91187113.  
 RA HSU I.C., METCALF R.A., SUN T., WELSH J.A., WANG N.J., HARRIS C.C.;  
 RL NATURE 350:427-428(1991).  
 RN [25]  
 RP VARIANT HEPATOCELLULAR CARCINOMAS MUTATIONS IN SOUTH AFRICA.  
 RX MEDLINE: 91187114.  
 RA BRESAC B., KEM M., WANDS J., OZTURK M.;  
 RL NATURE 350:429-431(1991).  
 RN [26]  
 RP VARIANT IN ANOGENITAL CARCINOMAS.  
 RX MEDLINE: 931010989.  
 RA CROOK T., VOUSDEN K.H.;  
 RL EMBO J. 11:3935-3940(1992).  
 RN [27]  
 RP VARIANT WITH DUPLICATION OF PRO-177-HIS-178.  
 RX MEDLINE: 93265016.

RA BHATIA K., GUTIERREZ M.I., MAGRATH I.T.;  
 RL HUM. MOL. GENET. 1:207-208(1992).  
 RN [28]  
 RP VARIANT IN BURKITT'S LYMPHOMAS.  
 RX MEDLINE: 93064692.  
 RA DUTHU A., DEBUIRE B., ROMANO J.W., EHRHART J.C., FISCELLA M., MAY E.,  
 RA APPELLA E., MAY P.;  
 RL ONCOGENE 7:2161-2167(1992).  
 RN [29]  
 RP VARIANT NASOPHARYNGEAL CARCINOMA THR-280.  
 RX MEDLINE: 93335329.  
 RA SUN Y., HEGAMER G., HENG Y.-J., HILDSHEIM A., CHEN J.-Y., CAO Y.,  
 RA YAO K.-T., COLEBURN N.H.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:6516-6520(1992).  
 RN [30]  
 RP VARIANT IN COLON TUMORS.  
 RX MEDLINE: 93330562.  
 RA HAMELIN R., JEGO N., LAURENT-PUIG P., VIDAUD M., THOMAS G.;  
 RL ONCOGENE 8:2213-2220(1993).  
 RN [31]  
 RP CHARACTERIZATION OF VARIANT ALA-143.  
 RX MEDLINE: 94283378.  
 RA ZHANG W., GUO X.-Y., HU G.-Y., LIU W.-B., SHAY J.W., DEISSEROTH A.B.,  
 RL EMBO J. 13:2535-2544(1994).  
 RN [32]  
 RP VARIANT LFS HIS-175; ARG-193; GLN-248; CYS-273 AND TYR-275.  
 RX MEDLINE: 95193787.  
 RA FREIBURG T., BARBIER N., YAN Y.-X., GARBER J.E., DREYFUS M.,  
 RA FRAMMENTI J.F. JR., LI F.P., FRIEND S.H.;  
 RL AM. J. HUM. GENET. 56:608-615(1995).  
 RN [33]  
 RP VARIANT LFS HIS-175.  
 RX MEDLINE: 96423319.  
 RA VARLEY J.M., MCGROWN G., THORNCROFT M., TRICKER K.J., TEARE M.D.,  
 RA SANTIBANEZ-KOREF M.F., HOULSTON R.S., MARTIN J., BIRCH J.M.,  
 RA EVANS D.G.R.;  
 RL J. MED. GENET. 32:942-945(1995).  
 RN [34]  
 RP VARIANT BA PHE-176; SER-245; TRP-248; TRP-282 AND GLN-286.  
 RX MEDLINE: 96233927.  
 RA AUDREZET M.-P., ROBASKIEWICZ M., MERCIER B., NOUSBAUD J.-B.,  
 RA HARDY E., BAILL J.-P., VOLANT A., LOZAC'H P., GOUEROU H., FEREC C.;  
 RL HUM. MUTAT. 7:109-113(1996).  
 RL HUM. MUTAT. 7:109-113(1996).

Note: remainder of annotations omitted.

| Query Match                   | Score           | DB 1: Length                    |
|-------------------------------|-----------------|---------------------------------|
| Best Local Similarity 100.0%; | 64;             | 393;                            |
| Matches 9;                    | Conservative 0; | Mismatches 0; Indels 0; Gaps 0; |

DB 315 SPQPKKPL 323  
 1 SPQPKKPL 9

RESULT 4  
 ID P53-CERAE STANDARD; PRT; 393 AA.  
 AC P13481;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS CERCOPIHUS AETHIOPS (GREEN MONKEY) (GRIYET).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-LIVER;  
 RX MEDLINE: 90045967.  
 RA RIGAUDY P., ECKHARDT W.;  
 RL NUCLEIC ACIDS RES. 17:8375-8375(1989).  
 CC -I- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES

CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC  
CC EMBL: X16384; G22796; .  
CC PIR: S06594; S06594.  
CC HSSP: P04637; 10LG.  
CC PROSITE: PS00348; P53; 1.  
CC ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
CC NUCLEAR PROTEIN: PHOSPHORYLATION; APOPTOSIS.  
CC  
CC FT DOMAIN 1 68 ASP/GLU-RICH (ACIDIC).  
CC FT 81 150 HYDROPHOBIC.  
CC FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
CC INTERACTION WITH DNA.  
CC FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC FT MOD\_RES 392 PHOSPHORYLATION (BY SIMILARITY).  
CC SEQUENCE 393 AA; 43696 MW; BBE7DC62 CRC32;  
SO  
Query Match 100.0%; Score 64; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 9, 18e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 315 SPQPKKKPL 323  
QY 1 SPQPKKKPL 9  
RESULT 5.  
ID P53 SHEEP STANDARD: PRT: 382 AA.  
AC P51664; 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53  
OS OVIS ARIES (SHEEP).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
CC EUTHERIA; ARTIODACTYLA.  
CC [1]  
CC SEQUENCE FROM N.A.  
CC  
CC TISSUE-BLOOD.  
CC RX MEDLINE: 95352828.  
CC RA DEUDET F., KETTMANN R., BURNY A., WILLEMS L.;  
CC RL DNA SEQ. 5:255-259 (1995).  
CC  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC  
CC EMBL: X81705; G602357; .  
CC PROSITE: PS00348; P53; 1.  
CC ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
CC NUCLEAR PROTEIN: PHOSPHORYLATION; APOPTOSIS.  
CC  
CC FT DOMAIN 1 66 ASP/GLU-RICH (ACIDIC).  
CC FT DOMAIN 300 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC MOD\_RES 381 381 PHOSPHORYLATION (BY SIMILARITY).  
SO

SO SEQUENCE 382 AA; 42809 MW; 0CB99A00 CRC32;  
Query Match 87.5%; Score 56; DB 1; Length 382;  
Best Local Similarity 88.9%; Pred. No. 1, 23e-02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
DB 304 SPQPKKKPL 312  
QY 1 SPQPKKKPL 9  
RESULT 6  
ID P53 RABBIT STANDARD: PRT: 391 AA.  
AC 095330; 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53  
OS ORYCTOLAGUS CUNICULUS (RABBIT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
CC EUTHERIA; LAGOMORPHA.  
CC [1]  
CC SEQUENCE FROM N.A.  
CC RC STRAIN-NEW ZEALAND;  
CC RX MEDLINE: 97208869.  
CC LE GOAS F., MAY P., RONCO P., CARON DE FROMENTEL C.;  
CC RL GENE 185:169-173 (1997).  
CC  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION (BY SIMILARITY).  
CC  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC  
CC EMBL: X90592; E194962; .  
CC DR PROSITE: PS00348; P53; 1.  
CC KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
CC NUCLEAR PROTEIN: PHOSPHORYLATION; APOPTOSIS.  
CC FT DOMAIN 1 70 ASP/GLU-RICH (ACIDIC).  
CC FT DOMAIN 308 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC FT MOD\_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).  
CC SEQUENCE 391 AA; 43435 MW; 30A36172 CRC32;  
SO  
Query Match 87.5%; Score 56; DB 1; Length 391;  
Best Local Similarity 88.9%; Pred. No. 1, 23e-02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
DB 313 SPQPKKKPL 321  
QY 1 SPQPKKKPL 9  
RESULT 7  
ID P53 RAT STANDARD: PRT: 391 AA.  
AC P10361; 009168; 01-MAR-1989 (REL. 10, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
CC EUTHERIA; RODENTIA.  
CC [1]



RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89083585.  
 RA SOUSST T.;  
 RN NUCLEIC ACIDS RES. 16:11384-11384(1988).  
 (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 93181268.  
 RA HULLA J.E., SCHNEIDER R.P.;  
 RL NUCLEIC ACIDS RES. 21:713-717(1993).  
 (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RA MATHUPALA S.P.;  
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC  
 DR EMBL: X13058; G56829; -;  
 DR EMBL: L07910; G205952; -;  
 DR EMBL: L07904; G205952; JOINED.  
 DR EMBL: L07905; G205952; JOINED.  
 DR EMBL: L07906; G205952; JOINED.  
 DR EMBL: L07907; G205952; JOINED.  
 DR EMBL: L07908; G205952; JOINED.  
 DR EMBL: L07909; G205952; JOINED.  
 DR EMBL: U90328; G1938365; -;  
 DR PIR: S02192; S02192; -;  
 DR HSSP: P04637; 1PES.  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 76 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 77 151 HYDROPHOBIC.  
 FT DOMAIN 277 391 HIGHLY BASIC AND MAY BE INVOLVED IN  
 FT INTERACTION WITH DNA.  
 FT DOMAIN 309 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).  
 FT VARIANT 103 103 G->S.  
 FT VARIANT 256 256 E->G.  
 FT CONFLICT 174 174 C->W (IN REF. 2).  
 SQ SEQUENCE 391 AA; E0114C18 CRC32;  
 Query Match 87.5%; Score 56; DB 1; Length 391;  
 Best Local Similarity 88.9%; Pred. No. 1,23e-02;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SYRIAN; TISSUE-KIDNEY;  
 RX MEDLINE: 92210007.  
 RA LEGROS Y., MCINTYRE P., SOUSST T.;  
 RL GENE 112:247-250(1992).  
 (2)  
 RP SEQUENCE FROM N.A.  
 RA HOU E.W., WISEMAN R.;  
 RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC  
 DR EMBL: M75144; G191415; -;  
 DR EMBL: U07182; G473579; -;  
 DR PIR: JH0633; JH0633.  
 DR HSSP: P04637; 1PES.  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 77 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 78 153 HYDROPHOBIC.  
 FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
 FT INTERACTION WITH DNA.  
 FT DOMAIN 314 326 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 395 395 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 188 188 G->S (IN REF. 2).  
 SQ SEQUENCE 396 AA; C2668ADE CRC32;  
 Query Match 82.8%; Score 53; DB 1; Length 396;  
 Best Local Similarity 77.8%; Pred. No. 7.05e-02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 318 SPORKRKL 326  
 Oy 1 SPORKRKL 9

RESULT 9  
 ID P53 HORSE STANDARD; PRT; 280 AA.  
 AC P79892; G29481;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN TP53 OR P53.  
 OS EOTUS CABALLUS (HORSE).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PERISSODACTYLA.  
 RN [1]  
 RP SEQUENCE OF 1-263 FROM N.A.  
 RC TISSUE-SPLEEN;  
 RX MEDLINE: 97070350.  
 RA PAZI K.A., KRAEGL S.A., GRIFFEY S.M., THEON A.P., MADEWELL B.R.;  
 RL CANCER LETT. 107:125-130(1996).  
 (2)  
 RP SEQUENCE OF 76-280 FROM N.A.  
 RX MEDLINE: 96293865.  
 RA NASIR L., REID S.W.;  
 RL DNA SEQ. 6:185-187(1996).



QY 1 111111  
1 SPOPKKPL 9

RESULT 11 STANDARD; PRT: 393 AA.  
ID P53-CTGR  
AC 009185; 064397; P97258; P97788;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS CRICETULUS GRISEUS (CHINESE HAMSTER).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHAUNG W., MI L.J., BOORSTEIN R.J.;  
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LIVER.  
RX MEDLINE; 97183659.  
RA LEE H., LARNER J.M., HAMLIN J.L.;  
RL GENE 184:177-183(1997).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND BCL-2 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
DR EMBL; Y08900; E303876; -  
DR EMBL; Y08901; E303863; -  
DR EMBL; U50395; G1842230; -  
DR PROSITE: PS00348; P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 1 74 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 75 150 HYDROPHOBIC.  
FT DOMAIN 316 390 HIGHLY BASIC AND MAY BE INVOLVED IN INTERACTION WITH DNA.  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
FT VARIANT 133 133 L -> Q (IN CELL LINE V79-4).  
FT VARIANT 135 135 C -> W (IN CELL LINE V79-4).  
FT CONFLICT 103 103 Y -> F (IN REF. 2).  
SQ SEQUENCE 393 AA; 43378 MW; 402EB149 CRC32.

Query Match 76.6%; Score 49; DB 1; Length 393;  
Best Local Similarity 77.8%; Pred. No. 6.63e-01;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 315 SPPKKKPL 323  
QY 1 111111  
1 SPOPKKPL 9

RESULT 12 STANDARD; PRT: 271 AA.  
ID RL7A-DROME  
AC P46223;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE 60S RIBOSOMAL PROTEIN L7A.

GN RPL7A.  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95257916.  
RA ARNES N., FRIED M.;  
RL MOL. CELL. BIOL. 15:2367-2373(1995).  
CC -1- SIMILARITY: BELONGS TO THE L7AE FAMILY OF RIBOSOMAL PROTEINS.  
DR EMBL; X82782; G80545; -  
DR FLYBASE; FBgn0014026; Rpl7A.  
DR PROSITE; PS01082; RIBOSOMAL\_L7AE; 1.  
KW RIBOSOMAL PROTEIN.  
SQ SEQUENCE 271 AA; 30677 MW; C804BEFE CRC32.

Query Match 73.4%; Score 47; DB 1; Length 271;  
Best Local Similarity 75.0%; Pred. No. 1.95e+00;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 6 PRPKKPL 13  
QY 2 PPKKPL 9

RESULT 13 STANDARD; PRT: 386 AA.  
ID P53-PELCA  
AC P41685;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS FELIS SILVESTRIS CATUS (CAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; CARNIVORA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LYMPH NODE.  
RX MEDLINE; 94333960.  
RA OKUDA M., UMEIDA A., SAKAI T., OHASHI T., MOMOI Y., YOUNG H.Y.,  
RA WATARI T., GOITSUKA R., TSUTSUMOTO H., HASEGAWA A.;  
RL INT. J. CANCER 58:602-607(1994).  
RN [2]  
RP SEQUENCE OF 34-354 FROM N.A.  
RX MEDLINE; 94114699.  
RA OKUDA M., UMEIDA A., MATSUMOTO Y., MOMOI Y., WATARI T., GOITSUKA R.,  
O'BRIEN S.J., TSUTSUMOTO H., HASEGAWA A.;  
RL J. VET. MED. SCI. 55:801-805(1993).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND BCL-2 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
DR EMBL; D26608; G538225; -  
DR EMBL; D16460; G575528; -  
DR PROSITE: PS00348; P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
FT CONFLICT 285 285 K -> R (IN REF. 2).  
SQ SEQUENCE 386 AA; 42692 MW; D6C7132A CRC32;

Query Match 73.4%; Score 47; DB 1; Length 386;

Best Local Similarity 66.7%; Pred. No. 1.95e+00;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB: 308 TPQKKKPL 316

QY 1 SPQKKKPL 9

|||||

QY 3 QPKKKPL 9

Search completed: Fri Sep 11 13:49:21 1998  
Job time : 8 secs.

RESULT 14  
ID: FTAL\_TRYBG STANDARD; PRT: 404 AA.

AC P15593:

DT 01-APR-1990 (REL. 14, CREATED)

DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)

DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)

DE RETROTRANSDUCIBLE ELEMENT SLACS 45 KD PROTEIN (ORF1).

OS TRYPAPOSOMA BRUCEI GAMBIENSE.

OC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA; KINETOPLASTIDA;

CC TRYPAPOSOMATIDAE.

RN [1]

RP SEQUENCE FROM N.A.

RA AKSOY S., WILLIAMS S., CHANG S., RICHARDS F.F.;

RL NUCLEIC ACIDS RES. 18:785-792(1990).

CC -1 SIMILARITY: WITH THE NUCLEIC ACID BINDING DOMAINS OF THE

CC RETROVIRAL GAG POLYPEPTIDES.

DR EMBL: X17078: G10534;

PIR: S14915; S14915.

KW TRANSDUCIBLE ELEMENT; DNA-BINDING; ZINC-FINGER; METAL-BINDING.

FT ZN\_FING 300 321 CYS-HIS MOTIF.

SO SEQUENCE 404 AA; 45463 MW; 5F648B65 CRC32;

Query Match 73.4%; Score 47; DB 1; Length 404;

Best Local Similarity 45.0%; Pred. No. 1.95e+00;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB: 167 TPQKKKPL 174

QY 1 SPQKKKPL 8

RESULT 15

ID FLIZ\_SALTY STANDARD; PRT: 50 AA.

AC P52628:

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE FLIZ PROTEIN (FRAGMENT).

GN FLIZ.

OS SALMONELLA TYPHIMURUM.

OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;

CC ENTEROBACTERIACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE: 90318310.

RA OHNISHI K., KITSUKAKE K., SUZUKI H., IINO T.;

RL MOL. GEN. GENET. 221:139-147(1990).

RN [2]

RP IDENTIFICATION.

RA RUD K.E.;

RL UNPUBLISHED OBSERVATIONS (JUN-1996).

CC -1 FUNCTION: MAY REGULATE SIGMA FACTOR ACTIVITY.

DR EMBL: X52624; NOT\_ANNOTATED\_CDS.

DR STIGENE; SGI???: FLIZ.

FT NON\_TER 50 50

SO SEQUENCE 50 AA; 5923 MW; DF74953F CRC32;

Query Match 71.9%; Score 46; DB 1; Length 50;

Best Local Similarity 71.4%; Pred. No. 3.30e+00;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB: 5 QPKRRPL 11

(TM)

Distribution rights by Oxford Molecular Ltd

protein - protein database search, using Smith-Waterman algorithm

```

Fri Sep 11 13:49:39 1998; Maspar time 3.68 seconds
1000000 words

```

generated.

1 SPQPKKKPL 9

2AM 150

140555 seqs, 42109429 residues

Existing first 45 summaries

sptremb16

1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mhc 6:sp\_organelle 7:sp\_phage 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp\_vertebrate  
13:sp\_unclassified

Mean 20.609; Variance 25.686; scale 0.802

ved by analysis of the total score distribution.

## SUMMARIES

|     |     |   |        |                        |          |
|-----|-----|---|--------|------------------------|----------|
| 5.0 | 305 | 1 | P78874 | FISSION YEAST (FRAGMEN | 3.24e+00 |
|-----|-----|---|--------|------------------------|----------|

|    |    |      |      |   |        |                        |          |
|----|----|------|------|---|--------|------------------------|----------|
| 45 | 44 | 68.8 | 1121 | 3 | 016866 | KINESIN LIKE PROTEIN A | 2.19e+01 |
|----|----|------|------|---|--------|------------------------|----------|

## ALIGNMENTS

|                               |   |                     |           |             |
|-------------------------------|---|---------------------|-----------|-------------|
| ID                            | 036006;   | PRELIMINARY;        | PRT;      | 391 AA.     |
| AD                            | 036006;   |                     |           |             |
| DT                            | 01-JAN-1998 (TREMBLREL. 05. CREATED)                              |                     |           |             |
| DT                            | 01-JAN-1998 (TREMBLREL. 05. LAST SEQUENCE UPDATE)                 |                     |           |             |
| DT                            | 01-JAN-1998 (TREMBLREL. 05. LAST ANNOTATION UPDATE)               |                     |           |             |
| DE                            | CELLULAR TUMOR ANTIGEN P53.                                       |                     |           |             |
| GN                            | P53.  |                     |           |             |
| OS                            | MARWATA MONAX.  |                     |           |             |
| OG                            | PLASMID PT7BLUE (R).  |                     |           |             |
| OC                            | UNCLASSIFIED.   |                     |           |             |
| RP                            | [1]   |                     |           |             |
| RA                            | SEQUENCE FROM N.A.  |                     |           |             |
| RL                            | FITTELESON M.A., RANGANATHAN P.N., CLAYTON M.M., ZHANG S.M.;      |                     |           |             |
| CC                            | ONCOGENE 15:327-336(1997).  |                     |           |             |
| CC                            | -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT |                     |           |             |
| CC                            | PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL   |                     |           |             |
| CC                            | CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY |                     |           |             |
| CC                            | REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED |                     |           |             |
| CC                            | FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF   |                     |           |             |
| CC                            | CYCLIN-DEPENDENT KINASES (BY SIMILARITY).                         |                     |           |             |
| CC                            | -1- SUBCELLULAR LOCATION: NUCLEAR.                                |                     |           |             |
| DR                            | EMBL: AJ001022, E351287; -.                                       |                     |           |             |
| DR                            | PROSITE: P500348; P53; 1.   |                     |           |             |
| KW                            | PLASMIID; ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION;   |                     |           |             |
| KW                            | ACTIVATOR; NUCLEAR PROTEIN; PHOSPHORYLATION.                      |                     |           |             |
| SO                            | SEQUENCE 391 AA; 43468 MW; 95FAB8F2 CRC32.                        |                     |           |             |
| Query Match                   |   |                     |           |             |
| Best Local Similarity 100.0%; |   | Score 64;           | DB 13;    | Length 391; |
| Matches 9; Conservative       |   | Pred. NO. 6.33e-04; |           |             |
|                               |   | Mismatches 0;       | Indels 0; | Gaps 0      |
| Db                            | 313 SPQPKKKPL 321   |                     |           |             |
| Qy                            | 1 SPQPKKKPL 9   |                     |           |             |
| RESULT 2                      |   |                     |           |             |
| ID                            | Q16811  | PRELIMINARY;        | PRT;      | 393 AA.     |
| AC                            | Q16811;   |                     |           |             |
| DT                            | 01-NOV-1996 (TREMBLREL. 01. CREATED)                              |                     |           |             |
| DT                            | 01-NOV-1996 (TREMBLREL. 01. LAST SEQUENCE UPDATE)                 |                     |           |             |
| DT                            | 01-JAN-1998 (TREMBLREL. 05. LAST ANNOTATION UPDATE)               |                     |           |             |
| DE                            | CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).                            |                     |           |             |

```

OS      HOMO SAPIENS (HUMAN) .
OC      EUKARYOTA: METAZOA; CHORDATA: VERTEBRATA: TETRAPODA; MAMMALIA:
NC      EUTHERIA; PRIMATES.
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE; 85126934.
RA      MATIASSEWSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,
RL      BENCHIMOL S.;
      EMO J. 3:3257-3262(1984).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      MEDLINE; 87064416.
RA      LAMB P., CRAWFORD L.;
      MOL. CELL. BIOL. 6:1376-1385(1986).
CC      -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
      PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
      CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
      REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
      FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
      CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: NUCLEAR.
CC      -1- SUBCELLULAR LOCATION: NUCLEAR.
DR      EMBL; M13121: G386994; JOINED.
DR      EMBL; M13112: G386994; JOINED.
DR      EMBL; M13113: G386994; JOINED.
DR      EMBL; M13114: G386994; JOINED.
DR      EMBL; M13115: G386994; JOINED.
DR      EMBL; M13116: G386994; JOINED.
DR      EMBL; M13117: G386994; JOINED.
DR      EMBL; M13118: G386994; JOINED.
DR      EMBL; M13119: G386994; JOINED.
DR      EMBL; M13120: G386994; JOINED.
DR      PROSITE; PS00348; P53; 1.
KW      REPEAT: TUMOR ANTIGEN; ANTI-ONCOGENE; DNA-BINDING;
      TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN;
      PHOSPHORYLATION.
FT      NON TER 393
SO      SEQUENCE 393 AA; 43698 MW; 3EA71431 CRC32;

Query Match 100.0%; Score 64; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 6,536-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SP0PKKPL 323
QY 1 SP0PKKPL 9

RESULT 3
ID 016807 PRELIMINARY: PRT: 393 AA.
AC 016807;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA: METAZOA; CHORDATA: VERTEBRATA; TETRAPODA; MAMMALIA:
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., YOUSDEN K.H., CROOK T.;
      EMO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
      PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
      CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
      REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
      FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
      CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; X60011: G506435; -
DR PROSITE; PS00348; P53; 1.
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;

```

```

KW NUCLEAR PROTEIN; PHOSPHORYLATION.
FT VARIANT 193 193 R -> H.
FT NON_TER 393
SQ SEQUENCE 393 AA; 43731 MW; 279BC9CB CRC32;

Query Match 100.0%; Score 64; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 6.53e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323
QY 1 SPQPKKKPL 9

RESULT 4
ID Q16808 PRELIMINARY; PRT; 393 AA.
AC Q16808;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RL EMO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; X60018; G506449; -.
DR PROSITE; PS00348; P53; 1.
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW NUCLEAR PROTEIN; PHOSPHORYLATION.
FT VARIANT 163 163 H -> Y.
FT NON_TER 393 393
FT SEQUENCE 393 AA; 43627 MW; AFDBA9E3 CRC32;

Query Match 100.0%; Score 64; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 6.53e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323
QY 1 SPQPKKKPL 9

RESULT 5
ID Q16535 PRELIMINARY; PRT; 393 AA.
AC Q16535;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN P53.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RL EMO J. 10:2879-2887(1991).
DR EMBL; X60017; G506447; -.
DR EMBL; X60015; G506443; -.
FT VARIANT 248 248 Q -> R.

```

FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43684 MW: 239818A9 CRC32:  
 Query Match 100.0%; Score 64; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.53e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPOPKKKPL 323  
 ||||||||  
 QY 1 SPOPKKKPL 9

RESULT 6  
 ID 015086 PRELIMINARY; PRT; 393 AA.  
 AC 015086;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TREMBLERL. 01, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60013; G506439; -.  
 FT VARIANT 246 T -> M.  
 FT NON\_TER 393  
 SQ SEQUENCE 393 AA: 43682 MW: 943862A3 CRC32:

Query Match 100.0%; Score 64; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.53e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPOPKKKPL 323  
 ||||||||  
 QY 1 SPOPKKKPL 9

RESULT 7  
 ID 016810 PRELIMINARY; PRT; 393 AA.  
 AC 016810;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60013; G506439; -.  
 FT VARIANT 246 D -> V.  
 FT NON\_TER 393  
 SQ SEQUENCE 393 AA: 43714 MW: 5F914579 CRC32:

Query Match 100.0%; Score 64; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.53e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPOPKKKPL 323  
 ||||||||  
 QY 1 SPOPKKKPL 9

RESULT 9  
 ID 016809 PRELIMINARY; PRT; 393 AA.  
 AC 016809;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60013; G506439; -.  
 FT VARIANT 246 D -> V.  
 FT NON\_TER 393  
 SQ SEQUENCE 393 AA: 43714 MW: 5F914579 CRC32:

Query Match 100.0%; Score 64; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.53e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPOPKKKPL 323  
 ||||||||  
 QY 1 SPOPKKKPL 9

RESULT 8  
 ID 016848 PRELIMINARY; PRT; 393 AA.  
 AC 016848;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTER V.;  
 RL MOL. CELL. BIOL. 6:4650-4656(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: M14694; G339814; -.  
 DR PROSITE: P500348; P53; 1.  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ACTIVATOR.  
 SQ SEQUENCE 393 AA: 43723 MW: DA7D302F CRC32:

Query Match 100.0%; Score 64; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.53e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPOPKKKPL 323  
 ||||||||  
 QY 1 SPOPKKKPL 9

RESULT 9  
 ID 016809 PRELIMINARY; PRT; 393 AA.  
 AC 016809;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60013; G506439; -.  
 FT VARIANT 246 D -> V.  
 FT NON\_TER 393  
 SQ SEQUENCE 393 AA: 43714 MW: 5F914579 CRC32:

Query Match 100.0%; Score 64; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.53e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPOPKKKPL 323  
 ||||||||  
 QY 1 SPOPKKKPL 9

RESULT 8  
 ID 016848 PRELIMINARY; PRT; 393 AA.  
 AC 016848;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTER V.;  
 RL MOL. CELL. BIOL. 6:4650-4656(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: X60013; G506439; -.

DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 213 213 Q -> R.  
 FT NON\_TER 393  
 SQ SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;

Query Match 100.0%; Score 64; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.53e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPOPKKPL 323  
 QY 1 SPOPKKPL 9

RESULT 10  
 ID Q15088 PRELIMINARY; PRT; 393 AA.  
 AC Q15088;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60016; G506445; -  
 DR EMBL: X60016; G506445; -  
 FT VARIANT 238 238 Y -> C.  
 FT NON\_TER 393  
 SQ SEQUENCE 393 AA; 43713 MW; A01E1523 CRC32;

Query Match 100.0%; Score 64; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.53e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPOPKKPL 323  
 QY 1 SPOPKKPL 9

RESULT 11  
 ID Q15087 PRELIMINARY; PRT; 393 AA.  
 AC Q15087;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60016; G506441; -  
 DR EMBL: X60016; G506441; -  
 FT VARIANT 1237 237 I -> M.  
 FT NON\_TER 393  
 SQ SEQUENCE 393 AA; 43694 MW; 9BB81992 CRC32;

Query Match 100.0%; Score 64; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.53e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPOPKKPL 323  
 QY 1 SPOPKKPL 9

QY 1 SPOPKKPL 9

RESULT 12  
 ID Q08901 PRELIMINARY; PRT; 1058 AA.  
 AC Q08901;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DE 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
 DE MITOTIC CHECKPOINT PROTEIN KINASE.  
 GN BUB1.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TAYLOR S.S., MCKEON F.;  
 RL CELL 89:727-735(1997).  
 DR EMBL: AF002823; G2150136; -  
 SQ SEQUENCE 1058 AA; 119562 MW; 73A3AFAS CRC32;

Query Match 84.4%; Score 54; DB 10; Length 1058;  
 Best Local Similarity 87.5%; Pred. No. 1.54e-01;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 528 POPKKPL 535  
 QY 2 POPKKPL 9

RESULT 13  
 ID Q09007 PRELIMINARY; PRT; 1102 AA.  
 AC Q09007;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE PROTEIN KINASE (FRAGMENT).  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-C57BL/6J;  
 RA PANGILINAN F., SPENCER F.;  
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U89795; G2335138; -  
 FT NON\_TER 1  
 SQ SEQUENCE 1102 AA; 124321 MW; F69C965F CRC32;

Query Match 84.4%; Score 54; DB 10; Length 1102;  
 Best Local Similarity 87.5%; Pred. No. 1.54e-01;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 572 POPKKPL 579  
 QY 2 POPKKPL 9

RESULT 14  
 ID Q21691 PRELIMINARY; PRT; 1057 AA.  
 AC Q21691;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE SIMILAR TO C. ELEGANS PROTEIN C14B1.7.  
 GN R04A9.2.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEIA; RHABDITIDA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,



Sun Sep 13 10:56:15 1998

RA COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,  
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
 RA JOHNSTON L., JONES M., KERSHAM J., KIRSTEN J., LAISTER N.,  
 RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,  
 RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFKEN L., ROOPRA A.,  
 RA SAUNDERS D., SHOWNKEEN R., SWALDON N., SMITH A., SONNHAMER E.,  
 RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
 RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
 RA WILKINSON-SPROAT J., WOHLDMAN P.,  
 RL NATURE 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA GEISEL C.  
 RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA WATERSTON R.,  
 RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U41550; G118046;  
 SQ SEQUENCE 1057 AA; 119712 MW; 813220C5 CRC32;

Query Match 81.3%; Score 52; DB 3; Length 1057;  
 Best Local Similarity 66.7%; Pred. No. 4.34e-01;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 76 GPPDKKKPL 84  
 QY 1 SPQPKKKPL 9

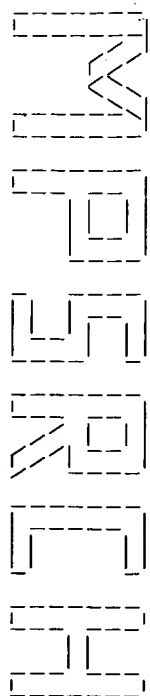
RESULT 15  
 ID 022777 PRELIMINARY; PRT; 528 AA.  
 AC 022777:  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE F4P9.2. PROTEIN.  
 DE GN F4P9.2.  
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;  
 OC CAPRIFALES; CRUCIFERAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA BOUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,  
 RA SYKES S.M., MASON T.M., KERLAVAGE A.R., ADAMS M.D., SOMERVILLE C.R.,  
 RA VENTER J.C.,  
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AC002332; G2459408;  
 SQ SEQUENCE 528 AA; 59760 MW; AF70E168 CRC32;

Query Match 79.7%; Score 51; DB 8; Length 528;  
 Best Local Similarity 66.7%; Pred. No. 7.23e-01;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 256 TSEPKKKPL 264  
 QY 1 SPQPKKKPL 9

Search completed: Fri, Sep 11 13:50:12 1998  
 Job time: 33 secs.

**THIS PAGE BLANK (USPTO)**



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsearch - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:44:08 1998; MasPar time 3.17 Seconds  
51.049 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-18  
Description: (1-10) from US08452843.pep  
Perfect Score: 68  
Sequence: 1 LPPOSTRKAL 10

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

a-geneseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 16.556; Variance 45.735; scale 0.362

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No.             |
|------------|-------|-------------|--------|-------|-------------|-----------------------|
| 1          | 68    | 100.0       | 25     | 24    | W14695      | Human p53 regulatory  |
| 2          | 68    | 100.0       | 319    | 24    | W28496      | Human p53 protein var |
| 3          | 68    | 100.0       | 319    | 24    | W28495      | Human p53 protein var |
| 4          | 68    | 100.0       | 335    | 24    | W28498      | Human p53 protein var |
| 5          | 68    | 100.0       | 335    | 24    | W28497      | Human p53 protein var |
| 6          | 68    | 100.0       | 353    | 24    | W28494      | Human p53 protein var |
| 7          | 68    | 100.0       | 353    | 24    | W28493      | Human p53 protein var |
| 8          | 68    | 100.0       | 353    | 24    | W13950      | Human p53 protein var |
| 9          | 68    | 100.0       | 361    | 21    | W13958      | Chimeric p53 protein  |
| 10         | 68    | 100.0       | 363    | 24    | W28479      | Human p53 protein var |
| 11         | 68    | 100.0       | 363    | 24    | W28480      | Human p53 protein var |
| 12         | 68    | 100.0       | 363    | 21    | W13971      | Modified p53 variant  |
| 13         | 68    | 100.0       | 363    | 21    | W13874      | Modified p53 variant  |
| 14         | 68    | 100.0       | 363    | 21    | W13973      | Modified p53 variant  |
| 15         | 68    | 100.0       | 363    | 21    | W13972      | Modified p53 variant  |
| 16         | 68    | 100.0       | 363    | 21    | W13975      | Modified p53 variant  |
| 17         | 68    | 100.0       | 363    | 21    | W13977      | Modified p53 variant  |
| 18         | 68    | 100.0       | 370    | 21    | W13957      | Chimeric p53 protein  |

|    |    |       |     |    |        |                       |          |
|----|----|-------|-----|----|--------|-----------------------|----------|
| 19 | 68 | 100.0 | 374 | 24 | W28482 | Human p53 protein var | 2.48e-01 |
| 20 | 68 | 100.0 | 374 | 24 | W28481 | Human p53 protein var | 2.48e-01 |
| 21 | 68 | 100.0 | 381 | 24 | W28490 | Human p53 protein var | 2.48e-01 |
| 22 | 68 | 100.0 | 381 | 24 | W28489 | Human p53 protein var | 2.48e-01 |
| 23 | 68 | 100.0 | 393 | 24 | W25155 | Human p53 variant fou | 2.48e-01 |
| 24 | 68 | 100.0 | 393 | 22 | W13953 | T284K modified human  | 2.48e-01 |
| 25 | 68 | 100.0 | 393 | 22 | W13948 | Human wild-type p53 t | 2.48e-01 |
| 26 | 68 | 100.0 | 393 | 21 | W05345 | Human p53 mutant N239 | 2.48e-01 |
| 27 | 68 | 100.0 | 393 | 21 | W13951 | Human tumour-derived  | 2.48e-01 |
| 28 | 68 | 100.0 | 393 | 22 | W13949 | T284K modified human  | 2.48e-01 |
| 29 | 68 | 100.0 | 393 | 22 | W13979 | Human tumour-derived  | 2.48e-01 |
| 30 | 68 | 100.0 | 393 | 22 | W13952 | Human tumour-derived  | 2.48e-01 |
| 31 | 68 | 100.0 | 393 | 18 | W02617 | Human p53 tumour supp | 2.48e-01 |
| 32 | 68 | 100.0 | 393 | 18 | R91933 | Wild type p53 protein | 2.48e-01 |
| 33 | 68 | 100.0 | 393 | 21 | W05348 | Human p53 mutant R282 | 2.48e-01 |
| 34 | 68 | 100.0 | 393 | 21 | W05344 | Human p53             | 2.48e-01 |
| 35 | 68 | 100.0 | 393 | 21 | W13970 | Modified p53 variant  | 2.48e-01 |
| 36 | 68 | 100.0 | 393 | 21 | W13968 | Modified p53 variant  | 2.48e-01 |
| 37 | 68 | 100.0 | 401 | 24 | W28487 | Human p53 protein var | 2.48e-01 |
| 38 | 68 | 100.0 | 401 | 24 | W28488 | Human p53 protein var | 2.48e-01 |
| 39 | 68 | 100.0 | 402 | 21 | W13965 | Chimeric p53 protein  | 2.48e-01 |
| 40 | 68 | 100.0 | 404 | 21 | W13963 | Chimeric p53 protein  | 2.48e-01 |
| 41 | 68 | 100.0 | 406 | 21 | W13966 | Chimeric p53 protein  | 2.48e-01 |
| 42 | 68 | 100.0 | 411 | 21 | W13967 | Chimeric p53 protein  | 2.48e-01 |
| 43 | 68 | 100.0 | 533 | 23 | W19763 | p53-GM-CSF immunostim | 2.48e-01 |
| 44 | 68 | 100.0 | 535 | 24 | W28492 | Human p53 protein var | 2.48e-01 |
| 45 | 68 | 100.0 | 535 | 24 | W28491 | Human p53 protein var | 2.48e-01 |

#### ALIGNMENTS

RESULT 1  
ID W14695 standard; Peptide: 25 AA.  
AC W14695;  
DE 24-NOV-1997 (first entry)  
DI Human p53 regulatory domain I.  
KW Tumour suppressor protein; p53; cancer; hyperproliferation;  
OS therapy; mmtic; heat shock protein; Dnak.  
KW Homo sapiens.  
FH Key  
FT binding\_site 4..13  
FT /label= P4421  
FT modified\_site 23  
FT /label= Phosphorylation  
FT /note= "cdc2 phosphorylation site"  
FN W09114794-A1.  
PD 24-APR-1997.  
PF 21-OCT-1996; G02605.  
PR 20-OCT-1995; GB-021544.  
PI (UYDU-) UNIV DUNDEE.  
PI Hupp TR, Lane DP;  
DR WPI: 97-245111/22.  
PT Substance which activates sequence specific DNA binding activity of  
PT latent p53 - useful for treatment of cancer or other  
PT hyperproliferative disorders  
PS Disclosure; Fig 15; 68bp; English.  
CC This peptide corresponds to amino acid residues 293-317 in the  
CC C-terminal negative regulatory domain of human tumour suppressor  
CC protein p53, and comprises regulatory domain I of p53. It is  
CC separated from regulatory domain II (W14697) by a tetramerisation  
CC domain. A binding site for monoclonal antibody P4421, which  
CC activates p53 for DNA binding, is present in domain I. Regulatory  
CC domain II includes the Dnak binding site (see also W14694). Human  
CC p53. Substances that activate the DNA binding activity of latent  
CC p53 are useful in the treatment of cancer and other  
CC hyperproliferative disorders.  
SQ Sequence 25 AA:  
Query Match 100.0%; Score 68; DB 24; Length 25;  
Best Local Similarity 100.0%; Pred. No. 2.48e-01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 LPPGSTRKRAL 10
RESULT 2
ID W28496 standard: Protein; 319 AA.
AC W28496;
DE 25-NOV-1997 (first entry)
KW Human p53 protein variant 360-325H.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutetin;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
OS tumour suppression; apoptosis.
OS Homo sapiens.
OS Synthetic.
FT Key Location/Qualifiers
FT /note- "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
FT msc_difference 145
PD 06-FEB-1997.
PD 17-JUL-1996: F01111.
PR 19-JUL-1995: FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI: 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 38; Page -: 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the domain 325-360 of p53. The present sequence is that of a
CC specifically claimed p53 variant designated 360-325H and comprising
CC the 325-360 domain, amino acids 75-325 of human wild-type p53 (but with
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.
CC The p53 variants are more active and more stable tumour suppressors
CC and apoptosis-inducing agents than wild-type p53 and are active where
CC the wild-type protein is not, i.e. they are not inactivated by dominant
CC negative or oncogenic mutants, nor by other cellular proteins (because
CC the leucine zipper domain prevents formation of inactive mixed
CC oligomers).
CC (Note: this sequence does not appear in the specification and has
CC been produced by modifying the given sequence of variant 360-325).
SQ Sequence 319 AA;
Query Match 100.0%; Score 68; DB 24; Length 319;
Best Local Similarity 100.0%; Pred. No. 2,48e-01;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 263 lppgstkrall 272
OY 1 LPPGSTRKRAL 10
RESULT 3
ID W28495 standard: Protein; 319 AA.
AC W28495;
DE 25-NOV-1997 (first entry)
KW Human p53 protein variant 360-325 encoded by p53178.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutetin;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Homo sapiens.
OS Synthetic.
OS W09704092-A1.
PD 06-FEB-1997.
PD 17-JUL-1996: F01111.
PR 19-JUL-1995: FR-008729.
PA (RHON ) RHONE POULENC RORER SA.

```

```

PI Bracco L, Conseiller E;
DR WPI: 97-132633/12.
DR N-PSDB; T86223.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 38; Pages 92-94; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the domain 325-360 of p53. The present sequence is that of a
CC specifically claimed p53 variant designated 360-325 and comprising
CC the 325-360 domain, amino acids 75-325 of human wild-type p53 and a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
SQ Sequence 319 AA;
Query Match 100.0%; Score 68; DB 24; Length 319;
Best Local Similarity 100.0%; Pred. No. 2,48e-01;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 263 lppgstkrall 272
OY 1 LPPGSTRKRAL 10
RESULT 4
ID W28498 standard: Protein; 335 AA.
AC W28498;
DE 25-NOV-1997 (first entry)
KW Human p53 protein variant 360h-325H.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutetin;
KW substitution; replacement; transactivation; hinge region;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Homo sapiens.
OS Synthetic.
FH Key Location/Qualifiers
FH region 39..53
FT /label= hinge
FT msc_difference 161
FT /note- "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
FT W09704092-A1.
PD 06-FEB-1997.
PD 17-JUL-1996: F01111.
PR 19-JUL-1995: FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI: 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 39; Page -: 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the domain 325-360 of p53. The present sequence is that of a
CC specifically claimed p53 variant designated 360h-325H and comprising
CC the 325-360 domain, separated from amino acids 75-325 of human
CC wild-type p53 (but with Arg182 replaced by His) by a synthetic hinge
CC sequence (GlySer)3, and with a leucine zipper domain at the C-terminal.
CC The p53 variants are more active and more stable tumour suppressors
CC and apoptosis-inducing agents than wild-type p53 and are active where
CC the wild-type protein is not, i.e. they are not inactivated by dominant
CC negative or oncogenic mutants, nor by other cellular proteins (because
CC the leucine zipper domain prevents formation of inactive mixed

```

CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 360h-325).  
 SO Sequence 335 AA.

Query Match 100.0%; Score 68; DB 24; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 2,48e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 279 lppgstkrall 288  
 |||||  
 1 lppgstkrall 10

RESULT 5  
 ID W28497 standard; Protein: 335 AA.

AC W28497;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360h-325 encoded by PEC179.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutelin;  
 KW substitution; replacement; transactivation; hinge region;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 FH Key  
 FT region Location/Qualifiers  
 FT 39..53  
 FT /label= hinge  
 PN MO9704092-A1.  
 PF 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PF 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 PI WPI: 97-132633/12.  
 DR N-PSDB; T86222.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 37; Pages 94-95; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360h-325 and comprising  
 CC the 325-360 domain, separated from amino acids 75-325 of human  
 CC wild-type p53 by a synthetic hinge sequence (Gly4Ser)3, and with a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SO Sequence 335 AA.

Query Match 100.0%; Score 68; DB 24; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 2,48e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 279 lppgstkrall 288  
 |||||  
 1 lppgstkrall 10

RESULT 6  
 ID W28494 standard; Protein: 353 AA.

AC W28494;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 393-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutelin;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;

KW tumour suppression; apoptosis.

OS Homo sapiens.

OS Synthetic.

FH key Location/Qualifiers

FT misc-difference 179

FT /note= "Arg residue at position 182 of wild-type  
 FT p53 has been mutated to His"

PN MO9704092-A1.

PF 06-FEB-1997.

PF 17-JUL-1996; F01111.

PF 19-JUL-1995; FR-008729.

PA (RHON ) RHONE POULENC RORER SA.

PI Bracco L, Conseiller E;  
 PI WPI: 97-132633/12.

PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 37; Page -: 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-393 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 393-325H and comprising  
 CC the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 393-325).  
 SO Sequence 353 AA.

Query Match 100.0%; Score 68; DB 24; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 2,48e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 297 lppgstkrall 306  
 |||||  
 1 lppgstkrall 10

RESULT 7  
 ID W28493 standard; Protein: 353 AA.

AC W28493;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 393-325 encoded by PEC177.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutelin;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 PN MO9704092-A1.  
 PN 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PF 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 PI WPI: 97-132633/12.  
 DR N-PSDB; T86222.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 37; Pages 90-92; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-393 of p53. The present sequence is that of

domain prevents formation of inactive mixed oligomers).

CC domain prevents formation of inactive mixed oligomers).  
SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 24; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 2.48e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 307 lppgstkrall 316  
 |||||||||  
 QY 1 LPPGSTKRALL 10

RESULT 11  
 W28480 standard; Protein: 363 AA.  
 W28480;  
 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Chimeric - Homo sapiens.  
 OS Synthetic - Herpes simplex virus.  
 FH Synthetic.  
 FT Key Location/Qualifiers  
 FT misc\_difference 189  
 FT /note= "Arg residue at position 182 of wild-type  
 p53 has been mutated to His"

MO9704092-A1.  
 06-FEB-1997.  
 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseller E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 30: Page - : 13pp: French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-325H and comprising  
 CC the VP16 TD, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant V-325).  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 24; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 2.48e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 307 lppgstkrall 316  
 |||||||||  
 QY 1 LPPGSTKRALL 10

RESULT 12  
 W13971 standard; Protein: 363 AA.  
 W13971;  
 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53R284del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.

PN MO9710843-A1.  
 PD 27-MAR-1997.  
 PR 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazoneis TD;  
 DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1; 51-52; 82pp: English.  
 CC Modified p53 variant p53R284del364-393 (W13971) has a Thr284 to Arg  
 CC substn. (see also W13949) and a deletion of the C-terminal 30  
 CC amino acids. The R284K substitution, introduced by site-directed  
 CC mutagenesis of p53 DNA, provides a novel p53-DNA contact between a  
 CC phosphate of the DNA backbone and p53. The C-terminal deletion  
 CC permits in vitro DNA binding. The variant provides the means for  
 CC pharmacological rescue of p53 function in cancer patients. Other  
 CC modified p53 constructs (W13949-50, W13953-54, W13968-77) have also  
 CC been produced. Nucleic acids coding for modified p53 can be used  
 CC for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 2.48e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 lppgstkrall 308  
 |||||||||  
 QY 1 LPPGSTKRALL 10

RESULT 13  
 W13974 standard; Protein: 363 AA.  
 W13974;  
 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53H273del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN MO9710843-A1.  
 PD 27-MAR-1997.  
 PR 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazoneis TD;  
 DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1; 56-57; 82pp: English.  
 CC Modified p53 variant p53H273del364-393 (W13974) has the tumour-  
 CC derived histidine 273 mutation (see also W13952) and a deletion  
 CC of the C-terminal 30 amino acids of wild-type p53 (see also  
 CC W13948). H13273 is a Class I p53 tumour mutation that affects DNA  
 CC binding. The C-terminal deletion, introduced by site-directed  
 CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
 CC tumour mutant. This provides the means for pharmacological rescue  
 CC of p53 function in cancer patients. Other modified p53 constructs  
 CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
 CC acids coding for modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 2.48e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 lppgstkrall 308  
 |||||||||  
 QY 1 LPPGSTKRALL 10

RESULT 14

ID W13973 standard; Protein; 363 AA.  
 AC W13973;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53Q248R284del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 OS apoptosis; protein engineering; DNA binding.  
 PN MO9710843-A1.  
 PD 27-MAR-1997.  
 PE 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004602.  
 PR 21-AUG-1996; US-697221.  
 PA (W13973) W13973 INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 treatment of cancer.  
 PS Example 1; 54-56; 82pp; English.  
 CC Modified p53 variant p53Q248R284del364-393 (W13973) has the tumour-  
 derived Gln248 mutation (see also W13951), a Thr284 to Arg substn.  
 CC (see also W13949) and a deletion of the 30 C-terminal amino acids  
 of wild-type p53 (W13948). Gln248 is a Class I p53 tumour mutation  
 that affects DNA binding. The T284R substitution, introduced by  
 site-directed mutagenesis of p53 DNA, provides a novel p53-DNA  
 contact between a phosphate of the DNA backbone and p53, and  
 restores DNA binding. The C-terminal deletion permits in vitro  
 DNA binding. The construct provides the means for pharmacological  
 rescue of p53 function in cancer patients. Other modified p53  
 constructs (W13949-50, W13953-54, W13968-77) have also been  
 produced. Nucleic acids coding for modified p53 can be used for  
 cancer gene therapy.  
 SQ Sequence: 363 AA;

Query Match 100.0%; Score 68; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 2,48e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 lppgstkrall 308  
 |||||||||  
 QY 1 LPPGSTKRAL 10

RESULT 15  
 ID W13972 standard; Protein; 363 AA.  
 AC W13972;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53Q248del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 OS apoptosis; protein engineering; DNA binding.  
 PN MO9710843-A1.  
 PD 27-MAR-1997.  
 PE 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004602.  
 PR 21-AUG-1996; US-697221.  
 PA (W13972) W13972 INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 treatment of cancer.  
 PS Example 1; 53-54; 82pp; English.  
 CC Modified p53 variant p53Q248del364-393 (W13972) has the tumour-  
 derived glutamine 248 mutation (see also W13951) and a deletion  
 of the C-terminal 30 amino acids of wild-type p53 (see also  
 W13948). Gln248 is a Class I p53 tumour mutation that affects DNA  
 binding. The C-terminal deletion, introduced by site-directed  
 mutagenesis of p53 DNA, activates the DNA binding of the p53  
 tumour mutant. This provides the means for pharmacological rescue  
 of p53 function in cancer patients. Other modified p53 constructs  
 (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
 acids coding for modified p53 can be used for cancer gene therapy.  
 SQ Sequence: 363 AA;

Query Match 100.0%; Score 68; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 2,48e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 lppgstkrall 308  
 |||||||||  
 QY 1 LPPGSTKRAL 10

Search completed: Fri Sep 11 13:44:24 1998  
 Job time: 16 secs.



(TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

MPSrch\_dp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Sep 11 13:44:42 1998; Maspar time 3.44 Seconds  
Tabular output not generated. 106.091 Million cell updates/sec

Title: >US-08-452-843-18  
Description: (1-10) from US08452843.pep  
Perfect Score: 68  
Sequence: 1 LPPGSTRKAL 10

Scoring table: PAM 150  
Gap 15

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r56  
1:p1r1 2:p1r2 3:p1r3 4:p1r4 5:p1r1d

Statistics: Mean 22.191; Variance 27.188; scale 0.816

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description            | Pred. No. |
|------------|-------|-------------|--------|----|--------|------------------------|-----------|
| 1          | 68    | 100.0       | 393    | 1  | DNHU53 | cellular tumor antigen | 1.95e-04  |
| 2          | 68    | 100.0       | 393    | 2  | S06594 | cellular tumor antigen | 1.95e-04  |
| 3          | 65    | 95.6        | 381    | 2  | S38824 | cellular tumor antigen | 1.06e-03  |
| 4          | 65    | 95.6        | 390    | 1  | DNMS53 | cellular tumor antigen | 1.06e-03  |
| 5          | 65    | 95.6        | 391    | 2  | S02192 | cellular tumor antigen | 1.06e-03  |
| 6          | 65    | 95.6        | 391    | 2  | JC6193 | tumor suppressor p53   | 1.06e-03  |
| 7          | 54    | 79.4        | 393    | 2  | JC6176 | tumor suppressor p53   | 3.88e-01  |
| 8          | 53    | 77.9        | 396    | 2  | JH0633 | cellular tumor antigen | 3.88e-01  |
| 9          | 53    | 77.9        | 519    | 2  | S70581 | dihydropyrimidinase    | 6.44e-01  |
| 10         | 51    | 75.0        | 631    | 2  | I52257 | epistatin - mouse      | 1.75e+00  |
| 11         | 50    | 73.5        | 230    | 2  | I56979 | nitric-oxide synthase  | 2.85e+00  |
| 12         | 49    | 72.1        | 630    | 2  | A39344 | tumor-associated mucin | 4.62e+00  |
| 13         | 49    | 72.1        | 963    | 2  | S45167 | chitin synthase (PC 2  | 4.62e+00  |
| 14         | 49    | 72.1        | 1199   | 2  | A41939 | G protein-coupled glu  | 4.62e+00  |
| 15         | 49    | 72.1        | 2185   | 2  | S60200 | acetyl-CoA carboxylas  | 4.62e+00  |
| 16         | 48    | 70.6        | 139    | 2  | S78253 | ribosomal protein l13  | 7.45e+00  |
| 17         | 48    | 70.6        | 386    | 2  | S51648 | cellular tumor antigen | 7.45e+00  |
| 18         | 48    | 70.6        | 455    | 2  | B36916 | site-specific recombi  | 7.45e+00  |
| 19         | 48    | 70.6        | 523    | 1  | A41648 | protein-tyrosine-phos  | 7.45e+00  |
| 20         | 48    | 70.6        | 559    | 2  | I49444 | SH3 binding protein -  | 7.45e+00  |
| 21         | 47    | 69.1        | 323    | 2  | A40433 | prephycoene pyrophosp  | 1.19e+01  |
| 22         | 47    | 69.1        | 758    | 2  | S65169 | hypothetical protein   | 1.19e+01  |
| 23         | 46    | 67.6        | 370    | 5  | 2CHR   | chloromucronate cyclol | 1.89e+01  |

| 24 | 46 | 67.6 | 370  | 5 | 1CHRA  | chloromucronate cyclol                    | 1.89e+01 |
|----|----|------|------|---|--|---|----------|
| 25 | 46 | 67.6 | 370  | 5 | 1CHRA <td>chloromucronate cyclol<td>1.89e+01</td></td>   | chloromucronate cyclol <td>1.89e+01</td>  | 1.89e+01 |
| 26 | 46 | 67.6 | 370  | 5 | B35255 <td>chloromucronate cyclol<td>1.89e+01</td></td>  | chloromucronate cyclol <td>1.89e+01</td>  | 1.89e+01 |
| 27 | 46 | 67.6 | 344  | 2 | S06602 <td>modulin antigen - fruit<td>1.89e+01</td></td> | modulin antigen - fruit <td>1.89e+01</td> | 1.89e+01 |
| 28 | 46 | 67.6 | 2233 | 2 | S63347 <td>acetyl-CoA carboxylas<td>1.89e+01</td></td>   | acetyl-CoA carboxylas <td>1.89e+01</td>   | 1.89e+01 |
| 29 | 45 | 66.2 | 181  | 2 | I52731 <td>gene mwhl1 protein -<td>2.99e+01</td></td>    | gene mwhl1 protein - <td>2.99e+01</td>    | 2.99e+01 |
| 30 | 45 | 66.2 | 321  | 2 | S31711 <td>alternative oxidase -<td>2.99e+01</td></td>   | alternative oxidase - <td>2.99e+01</td>   | 2.99e+01 |
| 31 | 45 | 66.2 | 576  | 2 | S65001 <td>probable membrane pro<td>2.99e+01</td></td>   | probable membrane pro <td>2.99e+01</td>   | 2.99e+01 |
| 32 | 45 | 66.2 | 633  | 2 | A49722 <td>endoglin precursor -<td>2.99e+01</td></td>    | endoglin precursor - <td>2.99e+01</td>    | 2.99e+01 |
| 33 | 45 | 66.2 | 1139 | 2 | S61918 <td>protein kinase C (EC<td>2.99e+01</td></td>    | protein kinase C (EC <td>2.99e+01</td>    | 2.99e+01 |
| 34 | 45 | 66.2 | 1526 | 2 | JN0598 <td>DNA topoisomerase (Ar<td>2.99e+01</td></td>   | DNA topoisomerase (Ar <td>2.99e+01</td>   | 2.99e+01 |
| 35 | 45 | 66.2 | 3131 | 2 | S39842 <td>emulatin synthetase -<td>2.99e+01</td></td>   | emulatin synthetase - <td>2.99e+01</td>   | 2.99e+01 |
| 36 | 44 | 64.7 | 386  | 2 | S45569 <td>nuclear factor I-A6 -<td>4.68e+01</td></td>   | nuclear factor I-A6 - <td>4.68e+01</td>   | 4.68e+01 |
| 37 | 44 | 64.7 | 406  | 2 | A40529 <td>nuclear factor I-A6 -<td>4.68e+01</td></td>   | nuclear factor I-A6 - <td>4.68e+01</td>   | 4.68e+01 |
| 38 | 44 | 64.7 | 440  | 2 | A39413 <td>transposase tmpr homo<td>4.68e+01</td></td>   | transposase tmpr homo <td>4.68e+01</td>   | 4.68e+01 |
| 39 | 44 | 64.7 | 452  | 2 | A36596 <td>lipopolysaccharide-pr<td>4.68e+01</td></td>   | lipopolysaccharide-pr <td>4.68e+01</td>   | 4.68e+01 |
| 40 | 44 | 64.7 | 461  | 2 | S45568 <td>nuclear factor I (clo<td>4.68e+01</td></td>   | nuclear factor I (clo <td>4.68e+01</td>   | 4.68e+01 |
| 41 | 44 | 64.7 | 498  | 2 | S45567 <td>nuclear factor I-A5 -<td>4.68e+01</td></td>   | nuclear factor I-A5 - <td>4.68e+01</td>   | 4.68e+01 |
| 42 | 44 | 64.7 | 509  | 2 | S45565 <td>nuclear factor I-A4 -<td>4.68e+01</td></td>   | nuclear factor I-A4 - <td>4.68e+01</td>   | 4.68e+01 |
| 43 | 44 | 64.7 | 522  | 2 | S09396 <td>nuclear factor I-A1.1<td>4.68e+01</td></td>   | nuclear factor I-A1.1 <td>4.68e+01</td>   | 4.68e+01 |
| 44 | 44 | 64.7 | 1502 | 2 | RGBYH1 <td>nuclear factor I-A1 -<td>4.68e+01</td></td>   | nuclear factor I-A1 - <td>4.68e+01</td>   | 4.68e+01 |
| 45 | 44 | 64.7 | 2123 | 2 | S55089 <td>CYC1/CYP3 transcripti<td>4.68e+01</td></td>   | CYC1/CYP3 transcripti <td>4.68e+01</td>   | 4.68e+01 |
|    |    |      |      |   |  | HEA1 protein - yeast <td>4.68e+01</td>    | 4.68e+01 |

## ALIGNMENTS

| RESULT 1 | ENTRY   | ENTRY TITLE               | ALTERNATE NAMES                |
|----------|---|---------------------------|--------------------------------|
| 1        | DNHU53  | #type complete            |                                |
|          | cellular tumor antigen p53 - human  |                           |                                |
|          | cellular phosphoprotein p53; oncoprotein p53; transformation suppressor p53; tumor suppressor p53 |                           |                                |
|          | ORGANISM  | #formal_name Homo sapiens | #common_name man               |
|          | DATE  | 05-Oct-1988               | #sequence_revision 18-Nov-1994 |
|          |   | 18-Sep-1997               | #text_change                   |

## ACCESSIONS

|  |
|--|
| A52224; A43073; J70436; S40773; S42669; A22837; A55060; A25397; B25397; S42452; S42453; I38085; I38086; I38087; I38088; I38089; I38090; I38091; I38092; I38093; A44905; I58354; I78850; S60153 |
|--|

## REFERENCE

|                   |   |
|-------------------|---|
| #authors          | Lamb, P.; Crawford, L.                                      |
| #journal          | Mol. Cell. Biol. (1986) 6:1379-1385                         |
| #title            | Characterization of the human p53 gene.                     |
| #cross-references | MOTID:87064416  |
| #accession        | A25224  |
| #molecule_type    | DNA   |
| #residues         | 1-393   |
| #cross-references | EMBL:X01405; GB:M13121; GB:N00032; NID:q189460; P1D:g386994 |

## REFERENCE

|                   |   |
|-------------------|---|
| #authors          | J70436  |
| #journal          | Buchanan, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.; Georgiev, G.P.       |
| #title            | Gene (1988) 70:245-252  |
| #cross-references | A variation in the structure of the protein-coding region of the human p53 gene.    |
| #accession        | MOTID:89108008  |
| #molecule_type    | DNA   |
| #residues         | 1-393   |
| #cross-references | EMBL:M22898; NID:q189474; P1D:q189476   |
| #note             | this 72-Arg allele appears to be about 5 times more frequent than the 72-Pro allele |

accession J70436

molecule\_type DNA

residues 1-71, 'P', '73-393' #label BUZ  
cross-references EMBL:M22898; NID:q189474; P1D:q189476  
this 72-Pro allele was found in both normal and malignant cell lines

## REFERENCE

|                   |   |
|-------------------|---|
| #authors          | S40773  |
| #journal          | Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.    |
| #title            | submitted to the EMBL Data Library, August 1990 |
| #accession        | S40773  |
| #molecule_type    | DNA   |
| #residues         | 1-393   |
| #cross-references | EMBL:X54156; NID:q35213; P1D:g35214             |

REFERENCE  
#authors Maltshewski, G.; Lamb, P.; Pim, D.; Peacock, J.; Crawford,  
#journal L.; Benchmol, S.  
#title EMBO J. (1984) 3:3257-3262  
#accession S42669  
##molecule-type mRNA  
##residues 101-393 #label MKI  
##cross-references EMBL:X01405; NID:g35215; PID:g642241  
REFERENCE  
#authors Zakut-Houri, R.; Bienen-Tadmor, B.; Givol, D.; Oren, M.  
#journal EMBO J. (1985) 4:1251-1255  
#title Human p53 cellular tumor antigen: cDNA sequence and  
expression in COS cells.  
#cross-references MUID:85230577  
#accession A22837  
##molecule-type mRNA  
##residues 1-71, 'P', '73-393 #label ZAK  
##cross-references EMBL:X02465; EMBL:M60950; NID:g35209; PID:g35210  
REFERENCE  
#authors Harlow, E.; Williamson, N.M.; Ralston, R.; Heliman, D.M.;  
#journal Adams, T.E.  
#title Mol. Cell. Biol. (1985) 5:1601-1610  
#accession A55060  
##molecule-type mRNA  
##residues 1-71, 'P', '73-272, 'H', 274-393 #label HA3  
##cross-references GB:K03199; NID:g189478; PID:g189479  
##experimental-source clone pR4-2, cell line A431  
REFERENCE  
#authors Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.;  
#journal Arai, N.; Rotter, V.  
#title Mol. Cell. Biol. (1986) 6:4650-4656  
#accession A25397  
##molecule-type mRNA  
##residues 1-78, 'T', 80-393 #label HAR  
##cross-references EMBL:M14694; NID:g339813; PID:g339814  
##experimental-source clone p53-H-1, transformed hybridoma SV-80 cell  
line  
#accession B25397  
##molecule-type mRNA  
##residues 1-71, 'P', '73-78, 'T', 80-393 #label HA2  
##cross-references EMBL:M14695; NID:g339815; PID:g339816  
##experimental-source clone p53-H-19, transformed hybridoma SV-80 cell  
line  
REFERENCE  
#authors Maltshewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider,  
#journal J.; Crawford, L.V.  
#title Mol. Cell. Biol. (1987) 7:961-963  
#accession S42452  
##molecule-type mRNA; DNA  
##residues 66-71, 'P', '73-79 #label MK2  
##experimental-source clone lambda C113  
#note 72-Cys was also found, and appears to represent a  
polymorphism  
#accession S42453  
##molecule-type mRNA; DNA  
##residues 66-79 #label MAT  
##experimental-source clone 36K  
REFERENCE  
#authors Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vouden, K.H.;  
#journal Crook, T.  
#title EMBO J. (1991) 10:2879-2887  
#cross-references MUID:92007731  
#accession I38082  
#status translated from GB/EMBL/DBJ

##molecule-type mRNA  
##residues 1-189, 'L', '189-393 #label LALT'  
##cross-references EMBL:X60010; NID:g506432; PID:g506433  
#note deletion of a C nucleotide causes a frameshift at  
position 566  
#accession I38083  
#status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-192, 'R', 194-393 #label F02  
##cross-references EMBL:X60011; NID:g506434; PID:g506435  
#accession I38084  
#status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-393 #label F03  
##cross-references EMBL:X60012; NID:g506436; PID:g506437  
#accession I38085  
#status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-245, 'T', 247-393 #label F04  
##cross-references EMBL:X60013; NID:g506438; PID:g506439  
#accession I38086  
#status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-236, 'I', 238-393 #label F05  
##cross-references EMBL:X60014; NID:g506440; PID:g506441  
#accession I38087  
#status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-247, 'Q', 249-393 #label F06  
##cross-references EMBL:X60015; NID:g506442; PID:g506443  
#accession I38088  
#status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-71, 'P', '73-237, 'Y', 239-393 #label F07  
##cross-references EMBL:X60016; NID:g506444; PID:g506445  
#accession I38089  
#status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-247, 'Q', 249-393 #label F08  
##cross-references EMBL:X60017; NID:g506446; PID:g506447  
#accession I38090  
#status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-71, 'P', '73-162, 'H', 164-393 #label F09  
##cross-references EMBL:X60018; NID:g506448; PID:g506449  
#accession I38091  
#status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-212, 'Q', 214-393 #label F10  
##cross-references EMBL:X60019; NID:g506450; PID:g506451  
#accession I38092  
#status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-253, 'D', 255-393 #label F11  
##cross-references EMBL:X60020; NID:g506452; PID:g506453  
#note all sequences submitted to the EMBL/GenBank/DBJ  
databases June 1991  
REFERENCE  
#authors Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.  
#journal Nucleic Acids Res. (1991) 19:6977  
#title An Alu polymorphism intragenic to the TP53 gene.  
#cross-references MUID:92107726  
#accession I38093  
#status translated from GB/EMBL/DBJ  
##molecule-type DNA  
##residues 1-393 #label RE2  
##cross-references EMBL:X54156; NID:g35213; PID:g35214  
REFERENCE  
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;  
#journal Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.;  
#status translated from GB/EMBL/DBJ  
#accession I38082  
#status translated from GB/EMBL/DBJ

#title p53 gene mutations in gastric cancer metastases and in gastric cancer cell lines derived from metastases.  
#cross-references MUID:92034678  
#accession A44905  
#molecule-type DNA  
#residues 246-247, 'W', 249-250 ##label YAM  
#cross-references CB:S63157; NID:9237829; PID:9237830  
#note sequence extracted from NCBI backbone (NCBIN:63157,  
Note: remainder of annotations omitted.

Query Match 100.0%; Score 68; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.95e-04;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTRAL 308  
|||  
QY 1 LPPGSTRAL 10

RESULT 2  
ENTRY S06594 #type complete  
TITLE cellular tumor antigen p53 - green monkey  
ORGANISM #formal\_name Cercopithecus aethiops #common\_name green monkey, grivet  
DATE 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 08-Sep-1997

ACCESSIONS S06594  
REFERENCE S06594  
#authors Rigaudy, P.; Eckhart, W.  
#journal Nucleic Acids Res. (1989) 17:8375  
#title Nucleotide sequence of a cDNA encoding the monkey cellular phosphoprotein p53.  
#cross-references MUID:90045967  
#accession S06594  
#molecule-type mRNA  
#residues 1-393 ##label RIG

CLASSIFICATION #cross-references EMBL:X16384; NID:922795; PID:922796  
KEYWORDS #superfamily cellular tumor antigen p53  
apoptosis; cell division control; DNA binding; homotrimer; nucleus; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 176,179,238,242 #binding-site zinc (Cys, His, Cys, Cys) #status predicted

392 #binding-site phosphoryl-RNA (Ser) (covalent) #status predicted  
SUMMARY #length 393 #molecular-weight 43696 #checksum 4263

Query Match 100.0%; Score 68; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.95e-04;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTRAL 308  
|||  
QY 1 LPPGSTRAL 10

RESULT 3  
ENTRY S38824 #type complete  
TITLE cellular tumor antigen p53, alternative splice form - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 25-Oct-1996  
ACCESSIONS S38824; S35478  
REFERENCE S38822  
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.; Shokat, O.; Rotter, V.  
#journal Mol. Cell. Biol. (1986) 6:3222-3239  
#title Immunologically distinct p53 molecules generated by alternative splicing.  
#accession S38824  
#molecule-type mRNA

#residues 1-381 ##label ARA  
REFERENCE S35478  
#authors Han, K.A.; Kulez-Martin, M.F.  
#journal Nucleic Acids Res. (1992) 20:1979-1981  
#title Alternatively spliced p53 RNA in transformed and normal cells of different tissue types.  
#accession S35478  
#status nucleic acid sequence not shown; translation not shown  
#molecule-type mRNA  
#residues 1-381 ##label HAN  
#cross-references EMBL:M13874; NID:9200202; PID:9200203  
#note the nucleotide sequence was submitted to the EMBL Data Library, July 1988

COMMENT This sequence, produced by alternative splicing of the tenth intron, lacks the carboxyl-terminal sequence necessary for covalent attachment of RNA. The function of this minor splice form is not known.

CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS alternative splicing; phosphoprotein  
FEATURE 1-44 #domain transcription activation #status predicted  
16-26 #label TRA  
99-289 #region conserved region I  
108-121 #domain DNA-binding core #status predicted #label DBC  
114-139 #region L1 loop  
160-172 #region conserved region II  
168-178 #region conserved region III  
231-252 #region conserved region IV  
233-248 #region L3 loop  
267-283 #region conserved region V  
313-319 #region nuclear location signal  
319-357 #region tetramer association  
7,9,12,18,23,37 #binding-site phosphate (Ser) (covalent) #status predicted  
173,176,235,239 #binding-site zinc (Cys, His, Cys, Cys) #status predicted  
312 #binding-site phosphate (Ser) (covalent) (by cdcl2 kinase) #status predicted

SUMMARY #length 381 #molecular-weight 42498 #checksum 8703

Query Match 95.6%; Score 65; DB 2; Length 381;  
Best Local Similarity 90.0%; Pred. No. 1.06e-03;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 296 LPPGSTRAL 305  
|||  
QY 1 LPPGSTRAL 10

RESULT 4  
ENTRY DNMS53 #type complete  
TITLE cellular tumor antigen p53 - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 05-Sep-1997  
ACCESSIONS A22739; S06336; A02684; S38822; S38823; I48703  
REFERENCE A22739  
#authors Bierz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.  
#journal EMBO J. (1984) 3:2179-2183  
#cross-references MUID:85027173  
#accession A22739  
#molecule-type DNA  
#residues 1-134, 'V', 136-390 ##label BIE

REFERENCE S06336  
#authors Chumakov, P.M.  
#journal Bioorg. Khim. (1987) 13:1691-1694  
#title Primary structure of DNA complementary to murine oncoprotein p53 mRNA.  
#cross-references MUID:88221682  
#accession S06336  
#status not compared with conceptual translation

```

##molecule_type mRNA
##residues 1-134,'V',136-390 ##label CHU
REFERENCE
#authors Zakut-Houri, R.; Oren, M.; Blenz, B.; Lavie, V.; Hazum, S.;
Givol, D.
#journal Nature (1983) 306:594-597
#title A single gene and a pseudogene for the cellular tumour
antigen p53.
#cross-references MUID:84068204
#accession A02684
##molecule_type mRNA
##residues 1-159,'H',161-167,'G',169-233,'I',235-390 ##label ZAK
REFERENCE
#authors Araki, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohet, O.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:3232-3239
#title Immunologically distinct p53 molecules generated by
alternative splicing.
#accession S38822
##status preliminary
##molecule_type mRNA
##residues 1-390 ##label ARA
#cross-references EMBL:M13872; NID:g200198; PID:g200199
#accession S38823
##status preliminary
##molecule_type mRNA
##residues 1-167,'G',169-233,'I',235-390 ##label AR2
#cross-references EMBL:M13873
REFERENCE
#authors Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.
#journal Nucleic Acids Res. (1984) 12:5609-5626
#title Cloning and expression analysis of full length mouse cDNA
sequences encoding the transformation associated protein
p53.
#cross-references MUID:84272240
#accession I48703
##status preliminary: translated from GB/EMBL/DBJ
#molecule_type mRNA
##residues 1-47,'R',49-78,'QW',82-390 ##label RES
#cross-references EMBL:X00741; NID:g53570; PID:g53571
COMMENT This DNA-binding protein plays an essential role in the regulation
of cell division, as it is required for the transition from phase
G0 to G1 of the cell cycle.
COMMENT The tetramer association region may exhibit a beta-turn,
beta-sheet, beta-turn, alpha-helix motif.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
phosphoprotein; transcription regulation; tumor suppressor
zinc
FEATURE
1-44 #domain transcription activation #status predicted
#label TRR\
#region conserved region I\
#domain DNA-binding core #status predicted #label DBC\
#region I1 loop\
#region conserved region II\
#region I2 loop\
#region conserved region III\
#region conserved region IV\
#region I3 loop\
#region conserved region V\
#region nuclear location signal\
#region tetramer association\
#binding site phosphate (Ser) (covalent) #status
predicted\
#binding site zinc (Cys, His, Cys, Cys) #status
predicted\
#binding site phosphate (Ser) (covalent) (by cdcd
kinase) #status predicted
#binding site phosphoryl-RNA (Ser) (covalent) #status
predicted
#length 390 #molecular-weight 43458 #checksum 1260
SUMMARY

```

```

Query Match          95.6%; Score 65; DB 1; Length 390;
Best Local Similarity 90.0%; Pred. No. 1.06e-03;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0

Db      296 LPPGSARRAL 305
Qy      1 LPPGSTKRAL 10

RESULT 5
ENTRY #type complete
TITLE cellular tumor antigen p53 - rat
ALTERNATE_NAMES gene p53 protein; nuclear oncoprotein p53
ORGANISM #format_name Rattus norvegicus #common_name Norway rat
DATE 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change
08-Sep-1997

ACCESSIONS 502192; S41149
REFERENCE S02192
#authors Soussi, T.; de Fromental, C.C.; Breugnot, C.; May, E.
#journal Nucleic Acids Res. (1988) 16:11384
#title Nucleotide sequence of a cDNA encoding the rat p53 nuclear
#protein.
#cross-references MUID:89083585
#accession S02192
#molecule_type mRNA
#residues 1-391 #label SOU
#cross-references EMBL:X13058; NID:g56828; PID:g56829
REFERENCE S41149
#authors Hulla, J.E.; Schneider, R.P.
#journal Nucleic Acids Res. (1993) 21:713-717
#title Structure of the rat p53 tumor suppressor gene.
#accession S41149
#status preliminary; nucleic acid sequence not shown;
translation not shown

#molecule_type DNA
#residues 1-173; 'W', 175-391 ##label HUL
#cross-references EMBL:L07909
#note the nucleotide sequence was submitted to the EMBL Data
Library, December 1992

GENETICS 25/2; 32/3; 123/3; 185/1; 259/2; 305/1; 329/3; 365/2
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc

FEATURE
174,177,236,240 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted
390 #binding_site phosphoryl-RNA (ser) (covalent) #status
predicted

SUMMARY #length 391 #molecular_weight 43451 #checksum 7105

Query Match          95.6%; Score 65; DB 2; Length 391;
Best Local Similarity 90.0%; Pred. No. 1.06e-03;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db      297 LPPGSARRAL 306
Qy      1 LPPGSTKRAL 10

RESULT 6
ENTRY #type complete
TITLE tumor suppressor p53 - rabbit
ORGANISM #format_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
08-Sep-1997

ACCESSIONS JC6193
REFERENCE JC6193
#authors Le Goss, F.; May, P.; Ronco, P.; de Fromental, C.C.
#journal Gene (1997) 185:169-173
#title cDNA cloning and immunological characterization of rabbit

```

```

GENETICS ##experimental_source kidney, strain MPI
#gene p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
FEATURE
179,182,241,245 #binding_site zinc (Cys, His, Cys, Cys) #status
395 #predicted
#binding_site phosphoryl-RNA (Ser) (covalent) #status
395 #predicted
SUMMARY #length 396 #molecular_weight 43631 #checksum 6617
Query Match 79.4%; Score 54; DB 2; Length 396;
Best Local Similarity 80.0%; Pred. No. 3,88e-01;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 302 LPPKSRKAL 311
||| |::|||
QY 1 LPPGSTRKAL 10
RESULT 9
ENTRY #type complete
TITLE dihydropyrimidinase - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change
10-Sep-1997
ACCESSIONS S70581
REFERENCE S70581
#authors Matsuda, K.; Sakata, S.; Kaneko, M.; Hamaajima, N.; Nonaka,
M.; Sasaki, M.; Yamaki, N.
#journal Biochim. Biophys. Acta (1996) 1307:140-144
#title Molecular cloning and sequencing of a cDNA encoding
dihydropyrimidinase from the rat liver.
#accession S70581
#status preliminary
#molecule_type mRNA
##residues 1-519 #label MAT
SUMMARY ##cross-references EMBL: D65704; NID: g1378018; PID: d1010479; PID: g1378019
#length 519 #molecular_weight 56833 #checksum 6037
Query Match 77.9%; Score 53; DB 2; Length 519;
Best Local Similarity 70.0%; Pred. No. 6,44e-01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 40 LPPEDTSRGL 49
|||| |::|
QY 1 LPPGSTRKAL 10
RESULT 10
ENTRY #type complete
TITLE epistatin - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
25-Oct-1996
ACCESSIONS I52257; I65210
REFERENCE I52257
#authors Vos, H.L.; De Vries, Y.; Hilkens, J.
#journal Blochm. Biophys. Res. Commun. (1991) 181:121-130
#title The mouse epistatin (Muc1) gene and its promoter. Rapid
evolution of the repetitive domain in the protein.
#cross-references MIMD: 92068178
#accession I52257
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
##residues 1-631 #label RES
#cross-references GB: M77226; NID: g199835; PID: g199837
#accession I65210
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA

```

##residues 1-631 ##label RE2  
##cross-references GB:M84683; NID:g199842; PID:g199843  
GENETICS  
#gene Muc1  
#introns 20/1; 454/3; 472/2; 517/1; 557/3; 607/3  
SUMMARY #length 631 #molecular-weight 64636 #checksum 6763

Query Match 75.0%; Score 51; DB 2; Length 631;  
Best Local Similarity 77.8%; Pred. No. 1.75e+00;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 595 VPPGSTRK 603  
Oy 1 LPPOSTKRA 9

RESULT 11  
ENTRY I56979 #type fragment  
TITLE nitric-oxide synthase (EC 1.14.13.39) - rat (fragment)  
ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
DATE 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 31-Oct-1997

ACCESSIONS  
REFERENCE I56979  
#authors Mohaupt, M.G.; Elzie, J.L.; Ahn, K.Y.; Clapp, W.L.; Wilcox, C.S.; Kone, B.C.  
#journal Kidney Int. (1994) 46:653-665  
#title Differential expression and induction of mRNAs encoding two inducible nitric oxide synthases in rat kidney.  
#cross-references MUID:95089280  
#accession I56979  
#status preliminary; translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-230 ##label RES  
#cross-references EMBL:002534; NID:g408464; PID:g408465  
CLASSIFICATION #superfamily nitric-oxide synthase; flavodoxin homology  
KEYWORDS calmodulin binding; chromoprotein; FAD; flavoprotein; FMN; heme; iron; NADP; oxidoreductase

SUMMARY #length 230 #checksum 5579

Query Match 73.5%; Score 50; DB 2; Length 230;  
Best Local Similarity 60.0%; Pred. No. 2.85e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 217 LPPOSTKRA 226  
Oy 1 LPPOSTKRAL 10

RESULT 12  
ENTRY A39344 #type complete  
TITLE tumor-associated mucin (MUC1) homolog precursor - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 03-Apr-1992 #sequence\_revision 03-Apr-1992 #text\_change 23-Feb-1997

ACCESSIONS  
REFERENCE A39344  
#authors Spicer, A.P.; Parry, G.; Patton, S.; Gendler, S.J.  
#journal J. Biol. Chem. (1991) 266:15099-15109  
#title Molecular cloning and analysis of the mouse homologue of the tumor-associated mucin, MUC1, reveals conservation of potential O-glycosylation sites, transmembrane and cytoplasmic domains and a loss of minisatellite-like polymorphism.  
#cross-references MUID:91332029  
#accession A39344  
#status preliminary  
##molecule\_type DNA  
##residues 1-630 ##label SPI  
#cross-references GB:M64928  
KEYWORDS cytoskeleton; transmembrane protein  
SUMMARY #length 630 #molecular-weight 64622 #checksum 4588

Query Match 72.1%; Score 49; DB 2; Length 630;  
Best Local Similarity 87.5%; Pred. No. 4.62e+00;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 594 VPPGSTRK 601  
Oy 1 LPPOSTKR 8

RESULT 13  
ENTRY S45167 #type complete  
TITLE chitin synthase (EC 2.4.1.16) 2 - yeast (Saccharomyces cerevisiae)  
ALTERNATE\_NAMES chitin-UDP acetyl-glucosaminyl-transferase 2; protein YBR038w; protein YBR0407  
ORGANISM #formal\_name Saccharomyces cerevisiae  
DATE 17-May-1994 #sequence\_revision 09-Sep-1994 #text\_change 06-Feb-1998

ACCESSIONS  
REFERENCE S45167; S45896; A30922  
#authors Silverman, S.J.  
#journal Yeast (1989) 5:459-467  
#title Similar and different domains of chitin synthases 1 and 2 of S. cerevisiae: two isozymes with distinct functions.  
#accession S45167  
##molecule\_type DNA  
##residues 1-963 ##label STL  
#cross-references EMBL:M23865; NID:g171219; PID:g171220  
REFERENCE S45893  
#authors Andre, B.; Czapluch, C.; Helm, C.; Jauniaux, J.C.; Urrestarazu, A.; Vissers, S.  
#submission submitted to the Protein Sequence Database, August 1994  
#accession S43896  
##molecule\_type DNA  
##residues 1-963 ##label AND  
#cross-references EMBL:235907; NID:g536257; PID:g536258; MIPS:YBR038w

GENETICS  
#gene SGD:CHS2  
#cross-references SGD:S000242; MIPS:YBR038w  
#map\_position 2R  
FUNCTION #description catalyzes the alpha-1,4-glycosylation of chitin by UDP-N-acetyl-D-glucosamine producing elongated chitin and UDP  
KEYWORDS glycosyltransferase; hexosyltransferase; transmembrane protein

FEATURE  
424-440 #domain transmembrane #status predicted #label TM1  
644-660 #domain transmembrane #status predicted #label TM2  
677-698 #domain transmembrane #status predicted #label TM3  
708-732 #domain transmembrane #status predicted #label TM4  
744-761 #domain transmembrane #status predicted #label TM5  
780-796 #domain transmembrane #status predicted #label TM6  
874-897 #domain transmembrane #status predicted #label TM7  
902-924 #domain transmembrane #status predicted #label TM8  
SUMMARY #length 963 #molecular-weight 109881 #checksum 1849

Query Match 72.1%; Score 49; DB 2; Length 963;  
Best Local Similarity 70.0%; Pred. No. 4.62e+00;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 227 LPPOSTKRAL 236  
Oy 1 LPPOSTKRAL 10

RESULT 14  
ENTRY A41939 #type complete  
TITLE G protein-coupled glutamate receptor - rat  
ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
DATE 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998

ACCESSIONS  
A41939; S15362

Search completed: Fri Sep 11 13:45:05 1998  
Job time : 23 secs.

```
REFERENCE      A41939
#authors      Houamed, K.M.; Kuiper, J.L.; Gilbert, T.L.; Haldeman, B.A.;
#journal      O'Hara, P.J.; Mulvihill, E.R.; Ames, W.; Hagen, F.S.
#title        Science (1991) 252:1318-1321
#cross-references GB:M61099; NID:g397806; PID:g204460
#cross-references MUID:92022526
#accession    A41939
#status       preliminary; not compared with conceptual translation
#molecule_type nucleic acid
#residues     1-1199 #label HOU
#cross-references GB:M61099; NID:g397806; PID:g204460
#experimental_source cerebellum
#note         sequence extracted from NCBI backbone (NCBIP:60785)
REFERENCE      S15362
#authors      Masu, M.; Tanabe, Y.; Tsuchida, K.; Shigemoto, R.; Nakanishi,
#journal      S.
#title        Nature (1991) 349:760-765
#cross-references MUID:91156047
#accession    S15362
#status       preliminary
#molecule_type mRNA
#residues     1-1199 #label MAS
#cross-references EMBL:X57569; NID:g56646; PID:g56647
#protein-coupled receptor; transmembrane protein
SUMMARY        #length 1199 #molecular-weight 135235 #checksum 5211

Query Match    72.1%; Score 49; DB 2; Length 1199;
Best Local Similarity 60.0%; Pred. No. 4.62e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 147 LPPGRTKKPI 156
OY 1 LPPGSTKRAL 10

RESULT 15
ENTRY
TITLE        S60200 #type complete
               acetyl-CoA carboxylase (EC 6.4.1.2) - smut fungus (Ustilago
               maydis)
ORGANISM      #formal_name Ustilago maydis #common_name corn smut
               15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change
               09-Sep-1997
ACCESSIONS    S60200; S49991
REFERENCE     S60200
#authors      Bailey, A.; Keon, J.; Owen, J.; Hargreaves, J.
#journal      Mol. Gen. Genet. (1995) 249:191-201
#title        The ACC1 gene, encoding acetyl-CoA carboxylase, is essential
               for growth in Ustilago maydis.
#accession    S60200
#molecule_type DNA
#residues     1-2185 #label BAI
#cross-references EMBL:Z46886; NID:g600097; PID:g600098
GENETICS
#gene         ACC1
#introns      14/1
CLASSIFICATION #superfamily lipoyl/biotin-binding homology; biotin
               carboxylase homology
               biotin; ligase
KEYWORDS
FEATURE       #domain biotin carboxylase homology #label BCH\
41-548         #domain lipoyl/biotin-binding homology #label LPB\
675-747         #binding_site biotin (lys) (covalent) #status predicted
714             #length 2185 #molecular-weight 240029 #checksum 9283
SUMMARY        #length 2185 #molecular-weight 240029 #checksum 9283

Query Match    72.1%; Score 49; DB 2; Length 2185;
Best Local Similarity 66.7%; Pred. No. 4.62e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

DB 155 PPGSAMRSL 163
OY 2 PPGSTKRAL 10
```







Best Local Similarity 100.0%; Pred. No. 1.08e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 299 LPESTRAL 308  
OY 1 LPESTRAL 10

RESULT 2  
ID P53 HUMAN STANDARD; PRT; 393 AA.  
AC P04637;  
DE 13 AUG-1987 (REL. 05, CREATED)  
DI 01 MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01 NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
GN TP53.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN (1);  
RX MEDLINE: 85230577.  
RA ZAKUT-HOURI R., BIENZ-TADMOR B., GIVOL D., OREN M.;  
RN EMBO J. 4:1251-1255(1985).  
RN (2)  
RX MEDLINE: 87064416.  
RA LAMB P., CRAMFORD L.;  
RN MOL. CELL. BIOL. 6:1379-1385(1986).  
RN (3)  
RX MEDLINE: 85267676.  
RA HARLOW E., WILLIAMSON N.M., RAISTON R., HELFMAN D.M., ADAMS T.E.;  
RN MOL. CELL. BIOL. 5:1601-1610(1985).  
RN (4)  
RX TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
RN ROTTER V.;  
RN MOL. CELL. BIOL. 6:4650-4656(1986).  
RN (5)  
RX MEDLINE: 89108008.  
RA BUCHMAN V.L., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P.,  
RN GEORGIEV G.P.;  
RN GENE 70:245-252(1988).  
RN (6)  
RX MEDLINE: 85126934.  
RA MATLASHENSKI G., LAMB P., PIM D., PEACOCK J., CRAMFORD L.,  
RN BENCHIMOL S.;  
RN EMBO J. 3:3257-3262(1984).  
RN (7)  
RX MEDLINE: 90191730.  
RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
RN ONCOGENE 5:423-426(1990).  
RN (8)  
RX MEDLINE: 94294808.  
RA CLORE G.M., OMICHENSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
RN APPELLE E., GROENENBORH A.M.;  
RN SCIENCE 265:386-391(1994).  
RN (9)  
RX MEDLINE: 95292092.  
RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
RN NAT. STRUCT. BIOL. 1:877-890(1994).  
RN (10)  
RX MEDLINE: 94294806.  
RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
RN SCIENCE 265:346-355(1994).

RN (11)  
RX MEDLINE: 94090335.  
RA HARRIS C.C.;  
RN SCIENCE 262:1980-1981(1993).  
RN (12)  
RX MEDLINE: 91289156.  
RA HOULSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
RN SCIENCE 253:49-53(1991).  
RN (13)  
RX MEDLINE: 96271983.  
RA DE VRIES E.M.G., RICHE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
RN LIAO D., SOUSSI T., KOVACH J.S., SOMMER S.S.;  
RN HUM. MUTAT. 7:202-213(1996).  
RN (14)  
RX MEDLINE: 91153807.  
RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
RN HUM. GENET. 86:369-370(1991).  
RN (15)  
RX MEDLINE: 92034774.  
RA LAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
RN CANCER RES. 51:6385-6387(1991).  
RN (16)  
RX MEDLINE: 91057657.  
RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
RN KIM D.H., KASSEL J., GRKYA M.A., BISCHOFF F.Z., TAINSKY M.A.,  
RN FRIEND S.H.;  
RN SCIENCE 250:1233-1238(1990).  
RN (17)  
RX MEDLINE: 91080929.  
RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATNER W., CHANG E.H.;  
RN NATURE 348:747-749(1990).  
RN (18)  
RX MEDLINE: 92147883.  
RA FELIX C.A., NAV M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
RN POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
RN KNUDSEN T., MINNA J.D.;  
RN J. CLIN. INVEST. 89:640-647(1992).  
RN (19)  
RX MEDLINE: 92228023.  
RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
RN GEHARDT M.C., ANDERSEN T.I., BORRESSEN A.-L., LI F.P., GARBER J.,  
RN STRONG L.C.;  
RN NEW ENGL. J. MED. 326:1309-1315(1992).  
RN (20)  
RX MEDLINE: 90295284.  
RA BARTER J., TIGGO R., GANNON J., LANE D.P.;  
RN ONCOGENE 5:893-899(1990).  
RN (21)  
RX MEDLINE: 91017544.  
RA RODRIGUES N.R., ROMAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
RN GANNON J.V., LANE D.P.;  
RN PROC. NATL. ACAD. SCI. U.S.A. 87:7555-7559(1990).  
RN (22)  
RX MEDLINE: 91282784.  
RA ISHIOKA C., SATO T., GANOCH M., SUZUKI T., SHIBATA H., KANAMARU R.,  
RN WAKI A., YAMAZAKI T.;  
RN BIOCHEM. BIOPHYS. RES. COMMUN. 177:901-906(1991).  
RN (23)  
RX MEDLINE: 91330175.  
RA VARIANTS OESOPHAGUS TUMORS L-152; A-155; H-175; F-176 AND H-273.  
RN CASSON A.G., MUKHOPADHYAY T., CLEARY K.R., RO J.Y., LEVIN B.,

RA ROTH J.A.;  
 RL CANCER RES. 51:4495-4499(1991).  
 RN [24]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN CHINA.  
 RX MEDLINE: 91187113.  
 RA HSU I.C., METCALF R.A., SUN T., WELSH J.A., WANG N.J., HARRIS C.C.;  
 RL NATURE 350:427-428(1991).  
 RN [25]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN SOUTH AFRICA.  
 RX MEDLINE: 91187114.  
 RA BRESSAC B., KEW M., WANDS J., OZTURK M.;  
 RL NATURE 350:429-431(1991).  
 RN [26]  
 RP VARIANTS IN ANGENITAL CARCINOMAS.  
 RX MEDLINE: 93010989.  
 RA CROOK T., VOUSDEN K.H.;  
 RL EMBO J. 11:3935-3940(1992).  
 RN [27]  
 RP VARIANT WITH DUPLICATION OF PRO-177-HIS-178.  
 RX MEDLINE: 93265016.  
 RA BHATIA K., GUTIERREZ M.I., MAGRATH I.T.;  
 RL HUM. MOL. GENET. 1:207-208(1992).  
 RN [28]  
 RP VARIANTS IN BURKITT'S LYMPHOMAS.  
 RX MEDLINE: 93064692.  
 RA DUTTU A., DEBIRE B., ROMANO J.W., EHRHART J.C., FISCELLA M., MAY E.,  
 RL ONCOGENE 7:2161-2167(1992).  
 RN [29]  
 RP VARIANT NASOPHARYNGEAL CARCINOMA THR-280.  
 RX MEDLINE: 92335329.  
 RA SUN Y., HEAMER G., HENG Y.-J., HILDSHEIM A., CHEN J.-Y., CAO Y.,  
 RA YAO K.-T., COLEBURN N.H.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:6516-6520(1992).  
 RN [30]  
 RP VARIANTS IN COLON TUMORS.  
 RX MEDLINE: 93330562.  
 RA HAMELIN R., JEGO N., LAURENT-PUIG P., VIDAUD M., THOMAS G.;  
 RL ONCOGENE 8:2213-2220(1993).  
 RN [31]  
 RP CHARACTERIZATION OF VARIANT ALA-143.  
 RX MEDLINE: 94283378.  
 RA ZHANG W., GUO X.-Y., HU G.-Y., LIU W.-B., SHAY J.W., DEISSEROTH A.B.;  
 RL EMBO J. 13:2535-2544(1994).  
 RN [32]  
 RP VARIANTS LFS HIS-175; ARG-193; GLN-248; CYS-273 AND TYR-275.  
 RX MEDLINE: 95193787.  
 RA FREBOURG T., BARBIER N., YAN Y.-X., GARBER J.E., DREYFUS M.,  
 RA FRAMMENT J.F. JR., LI F.P., FRIEND S.H.;  
 RL AM. J. HUM. GENET. 56:608-615(1995).  
 RN [33]  
 RP VARIANT LFS HIS-175.  
 RX MEDLINE: 96423319.  
 RA VARLEY J.M., MCGROWN G., THORNCROFT M., TRICKER K.J., TEARE M.D.,  
 RA SANTIBANEZ-KOEF M.F., HOULSTON R.S., MARTIN J., BIRCH J.M.,  
 RA EVANS D.G.R.;  
 RL J. MED. GENET. 32:942-945(1995).  
 RN [34]  
 RP VARIANTS BA-PHE-176; SER-245; TRP-248; TRP-282 AND GLN-286.  
 RX MEDLINE: 96233927.  
 RA AUDEZERT M.-P., ROBASZKIEWICZ M., MERCIER B., NOUSBAUM J.-B.,  
 RA HARDY E., BAIL J.-P., VOLANT A., LOZAC'H P., GOUDOU H., FEREC C.;  
 RL HUM. MUTAT. 7:109-113(1996).

Note: remainder of annotations omitted.

Query Match 100.0%; Score 68; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. NO. 1.08e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 299 LPPGSTRAL 308  
 OY 1 LPPGSTRAL 10

RESULT 3  
 ID P53\_MOUSE STANDARD; PRT; 390 AA.  
 AC P02340;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR TRP53 OR P53.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 85027173.  
 RA BIENZ B., ZAKUT-HOURI R., GIOVOL D., OREN M.;  
 RL EMBO J. 3:2179-2183(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84068204.  
 RA ZAKUT-HOURI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIOVOL D.;  
 RL NATURE 306:594-597(1983).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84272240.  
 RA JENKINS J.R., RUDGE K., REDMOND S., WADE-EVANS A.;  
 RL NUCLEIC ACIDS RES. 12:5609-5626(1984).  
 RN [4]  
 RP SEQUENCE FROM N.A. (CLONES PCD53; P53-M1 AND P53-M8).  
 RX MEDLINE: 87064640.  
 RA ARAI N., NOMURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.,  
 RA KOTTER V.;  
 RL MOL. CELL. BIOL. 6:3232-3239(1986).  
 RN [5]  
 RP SEQUENCE OF 222-258 FROM N.A.  
 RX MEDLINE: 92115342.  
 RA BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BREKNER R.,  
 RA BALMAIN A.;  
 RL ONCOGENE 6:2363-2369(1991).  
 RN [6]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE: 86149247.  
 RA SAMAD A., ANDERSON C.W., CARROLL R.B.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:897-901(1986).  
 RN [7]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE: 91006019.  
 RA MEER D.W., SIMON S., KIRKANA U., ECKHART W.;  
 RL EMBO J. 9:3253-3260(1990).  
 RN [8]  
 RP FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND BCL-2 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: X00876; G871421;  
 DR EMBL: X00877; G871421; JOINED.  
 DR EMBL: X00878; G871421; JOINED.  
 DR EMBL: X00879; G871421; JOINED.  
 DR EMBL: X00880; G871421; JOINED.  
 DR EMBL: X00881; G871421; JOINED.  
 DR EMBL: X00882; G871421; JOINED.  
 DR EMBL: X00883; G871421; JOINED.

DR EMBL: X00884; G871421; JOINED.  
DR EMBL: X00885; G871421; JOINED.  
DR EMBL: K01700; G200205; -  
DR EMBL: X01237; G53576; -  
DR EMBL: X00741; G53571; -  
DR EMBL: M13872; G200199; -  
DR EMBL: M13873; G200201; -  
DR EMBL: M13874; G200203; ALT\_SEQ.  
DR EMBL: S77930; G243255; -  
DR PIR: A02684; DNMS53.  
DR PIR: A22739; A22739.  
DR PIR: S38822; S38822.  
DR HSSP: P04637; IPES.  
DR TRANSFAC: T01806; -  
DR MGD: MGI:98834; TRP53.  
DR PROSITE: PS00348; P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS; DISEASE MUTATION.  
FT DOMAIN 1 75 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 76 150 HYDROPHOBIC.  
FT DOMAIN 276 390 INTERACTION WITH DNA.  
FT DOMAIN 308 320 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 312 312 PHOSPHORYLATION.  
FT MOD\_RES 389 389 PHOSPHORYLATION (BY CK2).  
FT VARIANT 135 135 A -> V (CAN COOPERATE WITH AN ACTIVATED  
FT VARIANT 168 168 E -> G (IN CLONE P53-M1).  
FT CONFLICT 48 48 Q -> R (IN REF. 3).  
FT CONFLICT 79 81 PVA -> QW (IN REF. 3).  
SQ SEQUENCE 390 AA; 43458 MM; 8943DD93 CRC32;  
Query Match 95.6%; Score 65; DB 1; Length 390;  
Best Local Similarity 90.0%; Pred. No. 7.79e-05;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 296 LPPGSKRAL 305  
QY 1 LPPGSKRAL 10  
RESULT 4  
ID P53\_RABBIT STANDARD; PRT; 391 AA.  
AC 095330;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLDULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS ORYCTOLAGUS CUNICULUS (RABBIT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; LAGOMORPHA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NEW ZEALAND;  
RX MEDLINE: 97208869.  
RA LE GOAS F., MAY P., RONCO P., CARON DE FROMENTEL C.;  
RU GENE 185169-173(1997).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

DR EMBL: X90592; E194962; -  
DR PROSITE: PS00348; P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 1 70 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 308 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 391 AA; 43435 MM; 30A36172 CRC32;  
Query Match 95.6%; Score 65; DB 1; Length 391;  
Best Local Similarity 90.0%; Pred. No. 7.79e-05;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 296 LPPGSKRAL 305  
QY 1 LPPGSKRAL 10  
RESULT 5  
ID P53\_RAT STANDARD; PRT; 391 AA.  
AC P10361; 009168;  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLDULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 89083585.  
RA SOUSSOT T.;  
RU NUCLEIC ACIDS RES. 16:11384-11384(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 93181268.  
RA HULLA J.E., SCHNEIDER R.P.;  
RU NUCLEIC ACIDS RES. 21:713-717(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY;  
RA MATHUPALA S.P.;  
RU SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
DR EMBL: X13058; G56829; -  
DR EMBL: L07910; G205952; -  
DR EMBL: L07904; G205952; JOINED.  
DR EMBL: L07905; G205952; JOINED.  
DR EMBL: L07906; G205952; JOINED.  
DR EMBL: L07907; G205952; JOINED.  
DR EMBL: L07908; G205952; JOINED.  
DR EMBL: L07909; G205952; JOINED.  
DR EMBL: U90328; G1938365; -  
DR PIR: S02192; S02192.  
DR HSSP: P04637; IPES.  
DR PROSITE: PS00348; P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.

FT DOMAIN 1 76 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 77 151 HYDROPHOBIC.  
FT DOMAIN 277 391 HIGHLY BASIC AND MAY BE INVOLVED IN  
INTERACTION WITH DNA.  
FT DOMAIN 309 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RSS 390 390 PHOSPHORYLATION (BY SIMILARITY).  
FT VARIANT 103 103 G -> S.  
FT VARIANT 256 256 E -> G.  
FT CONFLICT 174 174 C -> W (IN REF. 2).  
SQ SEQUENCE 391 AA: 43451 MW: E014C18 CRC32:  
Query Match 95.6%; Score 65; DB 1; Length 391;  
Best Local Similarity 90.0%; Pred. No. 7,79e-05;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 297 LPPGSTRAL 306  
QY 1 LPPGSTRAL 10  
RESULT 6  
ID P53-CANEA STANDARD; PRT: 276 AA.  
AC Q29537;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53.  
OS CANIS FAMILIARIS (DOG).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; CARNIVORA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BEAGLE;  
RX KRAEGLER S.A., PAZZI K.A., MADEWELL B.R.;  
RA MEDLINE: 953233915.  
RL CANCER LEFT. 92:181-186(1995).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
DE EMBL: S77819; G1000577; -;  
DR PROSITE: PS00348; P53; 1.  
KM ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
GN NON\_TER 1 1  
FT DOMAIN <1 35 ASP/GLU-RICH (ACIDIC).  
FT NON\_TER 276 276  
SQ SEQUENCE 276 AA: 30466 MW: 8C97AE44 CRC32:  
Query Match 88.2%; Score 60; DB 1; Length 276;  
Best Local Similarity 100.0%; Pred. No. 1,89e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 264 PPGSTRAL 272  
QY 2 PPGSTRAL 10  
RESULT 7  
ID P53-SPEBE STANDARD; PRT: 314 AA.  
AC Q64662;

DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53.  
OS SPERMOPHYTE BEECHYI (BEECHY GROUND SQUIRREL).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-THYMUS;  
RX MEDLINE: 95007566.  
RA RIVKINA M.B., CULLEN J.M., ROBINSON W.S., MARION P.L.;  
RL CANCER RES. 54:5430-5437(1994).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
DE EMBL: U43902; G1165312; -;  
DR PROSITE: PS00348; P53; 1.  
KM ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
GN NON\_TER 1 1  
FT DOMAIN 289 301 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT NON\_TER 314 314  
SQ SEQUENCE 314 AA: D07F433B CRC32:  
Query Match 88.2%; Score 60; DB 1; Length 314;  
Best Local Similarity 100.0%; Pred. No. 1,89e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 278 PPGSTRAL 286  
QY 2 PPGSTRAL 10  
RESULT 8  
ID P53-SHEEP STANDARD; PRT: 382 AA.  
AC P51664;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS OVIS ARIES (SHEEP).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; ARTIODACTYLA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD;  
RX MEDLINE: 95352828.  
RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
RL DNA SEQ. 5:255-259(1995).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2

CC CC EXPRESSION. NUCLEAR.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: X81705; G602357; -  
 KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN: PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 66 ASP/GLU-RICH (ACIDIC).  
 FT MOD.RES 381 381 PHOSPHORYLATION (BY SIMILARITY).  
 SO SEQUENCE 382 AA: 42809 MW: 0CB999A00 CRC32;  
 Query Match 1 1 88.2% Score 60; DB 1; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 1.89e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 DB 289 PGSTKRAL 297  
 QY 2 PGSTKRAL 10

RESULT 9  
 ID P53\_FELCA STANDARD: PRT: 386 AA.  
 AC P41685;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS FELLS SILVERSTRIS CATUS (CAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; CARNIVORA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LYMPH NODE.  
 RC MEDLINE: P4333960.  
 RA OKUDA M.; UMEBA A.; SAKAI T.; OHASHI T.; MOMOI Y.; YOUN H.Y.;  
 RA WATARI T.; GOITSUKA R.; TSUJIMOTO H.; HASEGAMA A.;  
 RL INT. J. CANCER 58:602-607(1994).  
 RN [2]  
 RP SEQUENCE OF 34-354 FROM N.A.  
 RX MEDLINE: 9411699.  
 RA OKUDA M.; UMEBA A.; MATSUMOTO Y.; MOMOI Y.; WATARI T.; GOITSUKA R.;  
 RA O'BRIEN S.J.; TSUJIMOTO H.; HASEGAMA A.;  
 RL J. VET. MED. SCI. 55:801-805(1993)  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: D26608; G538225; -  
 DR EMBL: D16460; G575528; -  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN: PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
 FT MOD.RES 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD.RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 285 285 K -> R (IN REF. 2).  
 SO SEQUENCE 386 AA: 42692 MW: D6C7132A CRC32;

Query Match 88.2% Score 60; DB 1; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 1.89e-03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 293 PGSTKRAL 301  
 QY 2 PGSTKRAL 10

RESULT 10  
 ID P53\_CRIGR STANDARD: PRT: 393 AA.  
 AC 009185; G64397; P97258; P97788;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR P53.  
 OS CRITETULUS GRISEUS (CHINESE HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CHANG W.; MI L.J.; BOORSTEIN R.J.;  
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER.  
 RC MEDLINE: 97183659.  
 RA LEE H.; LARNER J.M.; HAMLIN J.L.;  
 RL GENE 184:177-183(1997).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: Y08900; E303876; -  
 DR EMBL: Y08901; E303863; -  
 DR PROSITE: PS00395; G1842230; -  
 KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN: PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 74 ASP/GLU-RICH (ACIDIC).  
 FT MOD.RES 75 150 HYDROPHOBIC.  
 FT DOMAIN 316 390 HIGHLY BASIC AND MAY BE INVOLVED IN  
 FT INTERACTION WITH DNA.  
 FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD.RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
 FT VARIANT 133 133 L -> Q (IN CELL LINE V79-4).  
 FT VARIANT 135 135 C -> W (IN CELL LINE V79-4).  
 FT CONFLICT 103 103 Y -> F (IN REF. 2).  
 SO SEQUENCE 393 AA: 43378 MW: 402EB149 CRC32;

Query Match 79.4% Score 54; DB 1; Length 393;  
 Best Local Similarity 80.0%; Pred. No. 7.19e-02;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 299 LPPKSKRAL 308  
 QY 1 LPPKSKRAL 10

RESULT 11  
 ID P53\_MESAU STANDARD: PRT: 396 AA.  
 AC 000366; P97276;  
 DT 01-DEC-1992 (REL. 24, CREATED)

|           |   |
|-----------|---|
| RL        | SUBMITTED (OCT-1993) TO EMBL/GENBANK/DBJ DATA BANKS   |
| CC        | -1- FUNCTION: THIS ENZYME PRODUCES NITRIC OXIDE (NO) WHICH ACCOUNTS FOR THE BIOLOGIC ACTIVITY OF ENDOTHELIUM-DERIVED RELAXING FACTOR (EDRF) WHICH IS IMPORTANT IN REGULATION OF VASOMOTOR TONE AND BLOOD FLOW BY INHIBITING SMOOTH MUSCLE CONTRACTION AND PLATELET AGGREGATION. |
| CC        | -1- CATALYTIC ACTIVITY: L-ARGININE + N NADPH + M O(2) - CITRULLINE + NITRIC OXIDE + N NADP(+)   |
| CC        | -1- COFACTOR: THIS FLAVOPROTEIN BINDS ONE MOLE EACH OF FAD AND FMN.   |
| CC        | -1- ENZYME REGULATION: STIMULATED BY CALCIUM/CALMODULIN (BY SIMILARITY).  |
| CC        | -1- SIMILARITY: STRONG, TO OTHER NOS ISOZYMES. ALSO TO CYTOCHROME P-450 REDUCTASE.  |
| DR        | EMBL; U02534; G408465; -  |
| KW        | OXIDOREDUCTASE; NADP; FAD; FMN.   |
| FT        | NON_TER 1   |
| FT        | NON_TER 230 230   |
| SO        | SEQUENCE 230 AA; 25228 MW; 56CABDE CRC32;   |
| Db        | 217 LPBGTVAL 226  |
| Qy        | 1 LPBGTVAL 10   |
| RESULT 13 |   |
| ID        | MUCL_MOUSE STANDARD; PRT; 630 AA.   |
| AC        | 002496;   |
| DT        | 01-JUN-1994 (REL. 29, CREATED)  |
| DT        | 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)   |
| DT        | 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)   |
| DE        | MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM1) (EPISILIN). MUC1 OR MUC-1.  |
| CN        | MUS MUSCULUS (MUSE).  |
| OS        | MUS MUSCULUS (MUSE).  |
| OC        | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; EUTHERIA; RODENTIA.  |
| RA        | SEQUENCE FROM N.A.  |
| RA        | MEDLINE; 91332029.  |
| RL        | SPICER A.P., PARRY G., PATTON S., GENDLER S.J.; J. BIOL. CHEM. 266:15099-15109(1991).   |
| RN        | [2]   |
| RP        | SEQUENCE FROM N.A.  |
| RX        | MEDLINE; 92068178.  |
| RA        | VOS H.L., DEVARIES Y., HIKENS J.; BIOCHEM. BIOPHYS. RES. COMMON. 181:121-130(1991).   |
| RL        | -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN CYTOSKELETON.   |
| CC        | -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXCLUSIVELY LOCATED IN THE APICAL DOMAIN OF THE PLASMA MEMBRANE OF HIGHLY POLARIZED EPITHELIAL CELLS.  |
| CC        | -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF EPITHELIAL TISSUES. ABERRANTLY EXPRESSED IN EPITHELIAL CARCINOMAS.  |
| CC        | -1- P.M.: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.   |
| DR        | EMBL; M84688; G199843; -  |
| DR        | EMBL; U16175; G608490; -  |
| DR        | EMBL; M65132; G199841; -  |
| DR        | EMBL; M64928; G199844; JOINED.  |
| DR        | EMBL; M77226; G199837; -  |
| DR        | PIR; A39344; A39344.  |
| KW        | MGI; 92731; MUCL.   |
| KW        | GLYCOPROTEIN; SIGNAL; CYTOSKELETON; ACTIN-BINDING; TRANSMEMBRANE; REPEAT.   |
| KW        | SIGNAL  |
| FT        | CHAIN 1 20  |
| FT        | CHAIN 21 630  |
| FT        | DOMAIN 21 530   |
| FT        | DOMAIN 531 561  |
| FT        | DOMAIN 562 630  |
| FT        | DOMAIN 442 366  |
| FT        | REPEAT 42 61  |
| FT        | REPEAT 1 1  |
| FT        | REPEAT 16 x 20 AA TANDEM APPROXIMATE REPEATS.   |

```

FT REPEAT 62 81 2.
FT REPEAT 82 101 3.
FT REPEAT 102 122 4 (APPROXIMATE).
FT REPEAT 123 143 5 (APPROXIMATE).
FT REPEAT 144 164 6 (APPROXIMATE).
FT REPEAT 165 184 7.
FT REPEAT 185 204 8.
FT REPEAT 205 225 9 (APPROXIMATE).
FT REPEAT 226 246 10 (APPROXIMATE).
FT REPEAT 247 256 11.
FT REPEAT 257 286 12.
FT REPEAT 287 306 13.
FT REPEAT 307 326 14.
FT REPEAT 327 346 15.
FT REPEAT 347 366 16.
FT CARBOHYD 125 125 POTENTIAL.
FT CARBOHYD 275 275 POTENTIAL.
FT CARBOHYD 302 302 POTENTIAL.
FT CARBOHYD 335 335 POTENTIAL.
FT CARBOHYD 355 355 POTENTIAL.
FT CARBOHYD 366 366 POTENTIAL.
FT CARBOHYD 408 408 POTENTIAL.
FT CARBOHYD 432 432 POTENTIAL.
FT CARBOHYD 449 449 POTENTIAL.
FT CARBOHYD 508 508 POTENTIAL.
FT CARBOHYD 120 120 P -> L (IN REF. 2).
FT CONFLICT 121 121 L -> S (IN REF. 2).
FT CONFLICT 138 139 AT -> PA (IN REF. 2).
FT CONFLICT 140 140 T -> TT (IN REF. 2).
FT CONFLICT 423 423 F -> S (IN REF. 2).
FT CONFLICT 506 506 S -> D (IN REF. 2).
FT CONFLICT 602 602 O -> S (IN REF. 2).
SQ SEQUENCE 630 AA: 64622 MW: 4345D24A CRC32;

Query Match 72.1%; Score 49; DB 1; Length 630;
Best Local Similarity 87.5%; Pred. No. 1.24e+00;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

DR EMBL: M23865; G171220; -.
DR EMBL: Z35907; G536258; -.
DR PIR: S45167; S45167.
DR PIR: A30922; A30922.
DR SGD: L0000330; CHS2.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; TRANSMEMBRANE; CELL WALL;
MULTIGENE FAMILY.
FT TRANSMEM 423 443 POTENTIAL.
FT TRANSMEM 644 664 POTENTIAL.
FT TRANSMEM 678 698 POTENTIAL.
FT TRANSMEM 712 732 POTENTIAL.
FT TRANSMEM 744 764 POTENTIAL.
FT TRANSMEM 776 796 POTENTIAL.
FT TRANSMEM 876 896 POTENTIAL.
FT TRANSMEM 906 926 POTENTIAL.
SQ SEQUENCE 963 AA: 109881 MW: A24A3C4D CRC32;

Query Match 72.1%; Score 49; DB 1; Length 963;
Best Local Similarity 70.0%; Pred. No. 1.24e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```



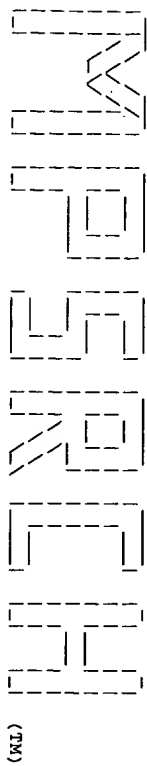
Page 9

CYTOLASMIC (POTENTIAL).  
 IV (POTENTIAL).  
 EXTRACELLULAR (POTENTIAL).  
 V (POTENTIAL).  
 CYTOLASMIC (POTENTIAL).  
 VI (POTENTIAL).  
 EXTRACELLULAR (POTENTIAL).  
 VII (POTENTIAL).  
 CYTOLASMIC (POTENTIAL).  
 GLYP/PRO-RICH.  
 GLN/PRO-RICH.  
 ASP/GLI-RICH (ACIDIC).  
 SER-RICH.  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.  
 NSNKSYSMSDEPGGAVPKG ->  
 KRRDPESPSPQCPSHAAQ (IN FORM BETA)  
 MISSING (IN FORM BETA).  
 NSNKSYSMSDEPGGAVPKG ->  
 KRRDPESPSPQCPSHAVQL (IN FORM BETA,  
 REF. 2).  
 P -> S (IN REF. 2).  
 NMA; 6ACDEFC3 CRC32:

```
Score 49;   DB 1;   Length 1194;
Pred.: No. 1.24e+00;
3; Mismatches 1; Indels 0; Gaps 0;
```

5:31 1998





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*

Run on: Fri Sep 11 13:45:49 1998; Maspar time 4.17 Seconds

Tabular output not generated. 101.074 Million cell updates/sec

Title: >US-08-452-843-18  
Description: (1-10) from US08452843.pep  
Perfect Score: 68  
Sequence: 1 LPPESTRKAL 10

Scoring table: PAM 150  
Gap 15

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mhc 6:sp\_organelle 7:sp\_phase 8:sp\_plant  
9:sp\_bacteria 10:sp\_fodent 11:sp\_virus 12:sp\_vertebrate  
13:sp\_unclassified

Statistics: Mean 22.568; Variance 23.537; scale 0.959

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description | Pred. No. |
|------------|-------|-------------|-----------|----|-------------|-----------|
| 1          | 68    | 100.0       | 393       | 2  | Q16809      | 1.62e-05  |
| 2          | 68    | 100.0       | 393       | 2  | Q16811      | 1.62e-05  |
| 3          | 68    | 100.0       | 393       | 2  | Q16807      | 1.62e-05  |
| 4          | 68    | 100.0       | 393       | 2  | Q16808      | 1.62e-05  |
| 5          | 68    | 100.0       | 393       | 2  | Q16535      | 1.62e-05  |
| 6          | 68    | 100.0       | 393       | 2  | Q15086      | 1.62e-05  |
| 7          | 68    | 100.0       | 393       | 2  | Q16810      | 1.62e-05  |
| 8          | 68    | 100.0       | 393       | 2  | Q16848      | 1.62e-05  |
| 9          | 68    | 100.0       | 393       | 2  | Q15087      | 1.62e-05  |
| 10         | 68    | 100.0       | 393       | 2  | Q15088      | 1.62e-05  |
| 11         | 60    | 88.2        | 281       | 4  | Q29475      | 1.62e-05  |
| 12         | 56    | 82.4        | 45        | 2  | Q09659      | 2.78e-03  |
| 13         | 55    | 80.9        | 648       | 2  | Q00392      | 5.84e-02  |
| 14         | 54    | 79.4        | 205       | 10 | Q35873      | 1.05e-01  |
| 15         | 53    | 77.9        | 238       | 11 | P89003      | 1.89e-01  |
| 16         | 53    | 77.9        | 286       | 11 | P89004      | 1.89e-01  |
| 17         | 53    | 77.9        | 286       | 11 | P90332      | 1.89e-01  |
| 18         | 53    | 77.9        | 378       | 11 | P89002      | 1.89e-01  |
| 19         | 53    | 77.9        | 519       | 10 | Q63150      | 1.89e-01  |
| 20         | 51    | 75.0        | 192       | 4  | Q28078      | 5.97e-01  |

|    |    |      |      |    |        |          |
|----|----|------|------|----|--------|----------|
| 21 | 51 | 75.0 | 193  | 10 | Q60551 | 5.97e-01 |
| 22 | 51 | 75.0 | 676  | 10 | Q60528 | 5.97e-01 |
| 23 | 50 | 73.5 | 285  | 4  | Q95326 | 1.05e+00 |
| 24 | 50 | 73.5 | 464  | 8  | Q04073 | 1.05e+00 |
| 25 | 49 | 72.1 | 2185 | 1  | Q12721 | 1.83e+00 |
| 26 | 49 | 72.1 | 2279 | 1  | P78820 | 1.83e+00 |
| 27 | 48 | 70.6 | 193  | 4  | Q28723 | 3.16e+00 |
| 28 | 48 | 70.6 | 193  | 10 | Q60408 | 3.16e+00 |
| 29 | 48 | 70.6 | 221  | 9  | Q65362 | 3.16e+00 |
| 30 | 48 | 70.6 | 391  | 13 | Q36006 | 3.16e+00 |
| 31 | 48 | 70.6 | 455  | 1  | Q53966 | 3.16e+00 |
| 32 | 48 | 70.6 | 558  | 1  | P87153 | 3.16e+00 |
| 33 | 48 | 70.6 | 602  | 4  | Q19115 | 3.16e+00 |
| 34 | 47 | 69.1 | 408  | 9  | P94947 | 5.42e+00 |
| 35 | 47 | 69.1 | 517  | 8  | Q23637 | 5.42e+00 |
| 36 | 47 | 69.1 | 758  | 1  | Q99299 | 5.42e+00 |
| 37 | 47 | 69.1 | 761  | 10 | P97321 | 5.42e+00 |
| 38 | 47 | 69.1 | 1323 | 10 | Q62645 | 5.42e+00 |
| 39 | 46 | 67.6 | 138  | 3  | Q15885 | 5.42e+00 |
| 40 | 46 | 67.6 | 154  | 9  | Q52244 | 9.22e+00 |
| 41 | 46 | 67.6 | 166  | 9  | Q45232 | 9.22e+00 |
| 42 | 46 | 67.6 | 660  | 9  | Q59218 | 9.22e+00 |
| 43 | 46 | 67.6 | 1095 | 2  | Q99458 | 9.22e+00 |
| 44 | 46 | 67.6 | 1999 | 2  | Q99940 | 9.22e+00 |
| 45 | 46 | 67.6 | 2003 | 2  | Q00306 | 9.22e+00 |

## ALIGNMENTS

| RESULT  | ID  | 1 | PRELIMINARY: | PRT: | 393 AA. |
|---|---|---|--------------|------|---------|
| AC  | Q16809  |   |              |      |         |
| DT  | 01-NOV-1996 (TREMBLREL. 01, CREATED)                              |   |              |      |         |
| DT  | 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)                 |   |              |      |         |
| DT  | 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)               |   |              |      |         |
| DE  | CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).                            |   |              |      |         |
| GN  | P53   |   |              |      |         |
| OS  | HOMO SAPIENS (HUMAN).   |   |              |      |         |
| OC  | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;    |   |              |      |         |
| OC  | EUTHERIA; PRIMATES.   |   |              |      |         |
| RN  | [1]   |   |              |      |         |
| RP  | SEQUENCE FROM N.A.  |   |              |      |         |
| RX  | MEDLINE: 92007731.  |   |              |      |         |
| RA  | FARELL P.J., ALIAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;       |   |              |      |         |
| RL  | EMBO J. 10:2879-2887(1991).                                       |   |              |      |         |
| CC  | -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT |   |              |      |         |
| CC  | PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL   |   |              |      |         |
| CC  | CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY |   |              |      |         |
| CC  | REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED |   |              |      |         |
| CC  | FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF   |   |              |      |         |
| CC  | CYCLIN-DEPENDENT KINASES (BY SIMILARITY).                         |   |              |      |         |
| CC  | -1- SUBCELLULAR LOCATION: NUCLEAR.                                |   |              |      |         |
| DR  | EMBL: X60019; G506451; -  |   |              |      |         |
| DR  | PROSITE: PS00348; P53; 1.   |   |              |      |         |
| KW  | ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  |   |              |      |         |
| KW  | NUCLEAR PROTEIN; PHOSPHORYLATION.                                 |   |              |      |         |
| FT  | VARIANT 213 213 Q -> R.   |   |              |      |         |
| FT  | NON_TER 213 213   |   |              |      |         |
| SO  | SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;                        |   |              |      |         |
| Query Match 100.0%; Score 68; DB 2; Length 393;             |   |   |              |      |         |
| Best Local Similarity 100.0%; Pred. No. 1.62e-05;           |   |   |              |      |         |
| Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |   |   |              |      |         |
| DB  | 299 LPPESTRKAL 308  |   |              |      |         |
| Qy  | 1 LPPESTRKAL 10   |   |              |      |         |
| RESULT 2  |   |   |              |      |         |
| ID  | Q16811  |   | PRELIMINARY: | PRT: | 393 AA. |
| AC  | Q16811  |   |              |      |         |
| DT  | 01-NOV-1996 (TREMBLREL. 01, CREATED)                              |   |              |      |         |

```

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
OC CELLULAR TUMOR ANTIGEN P53 (FRAGMENT)
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85126934.
RA BENCHMOL S.; LAMB P., PIM D., PEACOCK J., CRAWFORD L.,
RL EMO J. 3:3257-3262(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87064416.
RA LAMB P., CRAWFORD L.;
RL MOL. CELL. BIOL. 6:1379-1385(1986).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC EMBL: M13121: G386994; JOINED.
DR EMBL: M13112: G386994; JOINED.
DR EMBL: M13113: G386994; JOINED.
DR EMBL: M13114: G386994; JOINED.
DR EMBL: M13115: G386994; JOINED.
DR EMBL: M13116: G386994; JOINED.
DR EMBL: M13117: G386994; JOINED.
DR EMBL: M13118: G386994; JOINED.
DR EMBL: M13119: G386994; JOINED.
DR EMBL: M13120: G386994; JOINED.
DR PROSITE: P500348; P53; 1.
KW REPEAT: TUMOR ANTIGEN; ANTI-ONCOGENE; DNA-BINDING;
KW TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN;
KW PHOSPHORYLATION.
FT NON_TER 393
SQ SEQUENCE 393 AA: 3EA71431 CRC32:
Query Match 100.0%; Score 68; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.62e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 299 LPPGSTRAL 308
OY 1 LPPGSTRAL 10
RESULT 3 PRELIMINARY; PRT; 393 AA.
AC 016807;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT)
GN P53.
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RL EMO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.

```

```

DR EMBL: X60011: G506435; 1.
DR PROSITE: P500348; P53; 1.
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW NUCLEAR PROTEIN; PHOSPHORYLATION.
FT VARIANT 193 193 R -> H.
FT NON_TER 393 393
SQ SEQUENCE 393 AA: 43731 MW: 2798C9CB CRC32:
Query Match 100.0%; Score 68; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.62e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 299 LPPGSTRAL 308
OY 1 LPPGSTRAL 10
RESULT 4 PRELIMINARY; PRT; 393 AA.
AC 016808;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT)
GN P53.
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RL EMO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC EMBL: X60018: G506449; 1.
DR PROSITE: P500348; P53; 1.
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW NUCLEAR PROTEIN; PHOSPHORYLATION.
FT VARIANT 163 163 H -> Y.
FT NON_TER 393 393
SQ SEQUENCE 393 AA: 43627 MW: AFD8A9E3 CRC32:
Query Match 100.0%; Score 68; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.62e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 299 LPPGSTRAL 308
OY 1 LPPGSTRAL 10
RESULT 5 PRELIMINARY; PRT; 393 AA.
AC 016535;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT)
GN P53.
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RL EMO J. 10:2879-2887(1991).

```

DR EMBL: X60017: G506447: -  
 DR EMBL: X60015: G506443: -  
 FT VARIANT 248 248 Q -> R.  
 FT NON-TER 393 393  
 SQ SEQUENCE 393 AA: 43684 MW: 239818A9 CRC32:

Query Match  
 Best Local Similarity 100.0%; Score 68; DB 2; Length 393;  
 Pred. No. 1.62e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTRAL 308  
 QY 1 LPPGSTRAL 10

RESULT 6  
 ID Q15086 PRELIMINARY: PRT: 393 AA.  
 AC Q15086;  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA;  
 OC EUTHERIA: PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60013: G506439: -  
 DR EMBL: X60013: G506439: -  
 FT VARIANT 246 246 T -> M.  
 FT NON-TER 393 393  
 SQ SEQUENCE 393 AA: 43682 MW: 943B62A3 CRC32:

Query Match  
 Best Local Similarity 100.0%; Score 68; DB 2; Length 393;  
 Pred. No. 1.62e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTRAL 308  
 QY 1 LPPGSTRAL 10

RESULT 7  
 ID Q16810 PRELIMINARY: PRT: 393 AA.  
 AC Q16810;  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMELREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA;  
 OC EUTHERIA: PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60014: G506441: -  
 DR EMBL: X60014: G506441: -  
 FT VARIANT 237 237 I -> M.  
 SQ SEQUENCE 393 AA: 43694 MW: 98B81992 CRC32:

FT VARIANT 254 254 D -> V.  
 FT NON-TER 393 393  
 SQ SEQUENCE 393 AA: 43714 MW: 5F914579 CRC32:

Query Match  
 Best Local Similarity 100.0%; Score 68; DB 2; Length 393;  
 Pred. No. 1.62e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTRAL 308  
 QY 1 LPPGSTRAL 10

RESULT 8  
 ID Q16848 PRELIMINARY: PRT: 393 AA.  
 AC Q16848;  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMELREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA;  
 OC EUTHERIA: PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTER V.;  
 RL MOL. CELL. BIOL. 6:4650-4656(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: M14694: G339814: -  
 DR PROSITE: P500348; P53; 1.  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ANTI-ONCOGENE; DNA-BINDING;  
 KW TRANSCRIPTION REGULATION; ACTIVATOR.  
 SQ SEQUENCE 393 AA: 43723 MW: DA7D302F CRC32:

Query Match  
 Best Local Similarity 100.0%; Score 68; DB 2; Length 393;  
 Pred. No. 1.62e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTRAL 308  
 QY 1 LPPGSTRAL 10

RESULT 9  
 ID Q15087 PRELIMINARY: PRT: 393 AA.  
 AC Q15087;  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA;  
 OC EUTHERIA: PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60014: G506441: -  
 DR EMBL: X60014: G506441: -  
 FT VARIANT 237 237 I -> M.  
 FT NON-TER 393 393  
 SQ SEQUENCE 393 AA: 43694 MW: 98B81992 CRC32:

Query Match 100.0%; Score 68; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.62e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 299 LPPGSTRAL 308  
QY 1 LPPGSTRAL 10

RESULT 10  
ID 015088 PRELIMINARY; PRT; 393 AA.  
AC 015088;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
GN P53.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RL EMBL J. 10:2879-2887(1991).  
DR EMBL; X60016; G506445; -  
FT VARIANT 238 238 Y -> C.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA; 43713 MW; A01E1523 CRC32;

Query Match 100.0%; Score 68; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.62e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 299 LPPGSTRAL 308  
QY 1 LPPGSTRAL 10

RESULT 11  
ID 029475 PRELIMINARY; PRT; 281 AA.  
AC 029475;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN P53.  
OS CANIS FAMILIARIS (DOG).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; CARNIVORA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA TISSUE-MAMMARY GLAND;  
RA MEDLINE; 97194812.  
RA VAN LEEUWEN I., RUTEMAN G.R., HELLMEN E., CORNELISSE C.C.J.,  
RA DEVLIEP P.;  
RL ANTICANCER RES. 16:3737-3744(1996).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL; L37107; G1463021; -  
DR PROSITE; PS00348; P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
FT NON\_TER 1 1  
FT NON\_TER 281 281  
SQ SEQUENCE 281 AA; 31762 MW; FC7BAE31 CRC32;

Query Match 88.2%; Score 60; DB 4; Length 281.

Best Local Similarity 100.0%; Pred. No. 2.78e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 193 PPGSTRAL 201  
QY 2 PPGSTRAL 10

RESULT 12  
ID 099659 PRELIMINARY; PRT; 45 AA.  
AC 099659;  
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)  
DE CELLULAR PHOSPHOPROTEIN P53 (FRAGMENT).  
GN P53.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA FILIPPINI G., SOLDATI G.;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; U63714; G1753089; -  
FT NON\_TER 1 1  
FT NON\_TER 45 45  
SQ SEQUENCE 45 AA; 5170 MW; 09281164 CRC32;

Query Match 82.4%; Score 56; DB 2; Length 45;  
Best Local Similarity 100.0%; Pred. No. 3.21e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 38 LPPGSTR 45  
QY 1 LPPGSTR 8

RESULT 13  
ID 000392 PRELIMINARY; PRT; 648 AA.  
AC 000392;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE UV RADIATION RESISTANCE ASSOCIATED PROTEIN.  
GN UVRAG.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA PERELMAN B., DAFNI N., NAIMAN T., ELI D., YAKOV M., FENG T.L.Y.,  
RA SINHA S., WEBER G., KHODAEI S., SANCAR A., DOTAN I., CANAANI D.;  
RL GENOMICS 41:397-405(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA CANAANI D.;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA CANAANI D.;  
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; X99050; E354226; -  
SQ SEQUENCE 648 AA; 72363 MW; 7877028C CRC32;

Query Match 80.9%; Score 55; DB 2; Length 648;  
Best Local Similarity 80.0%; Pred. No. 5.84e-02;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 22 LPPGSTRAL 31  
QY 1 LPPGSTRAL 10

RESULT 14  
ID 035873 PRELIMINARY: PRT: 205 AA.  
AC 035873:  
DT 01-JAN-1998 (TREMBLERL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLERL. 05, LAST SEQUENCE UPDATE)  
DE 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN P53.  
OS CRICETULUS GRISEUS (CHINESE HAMSTER).  
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
OC EUTHERIA: RODENTIA.  
RN 11  
RP SEQUENCE FROM N.A.  
RA RAINALDI G., MARCHETTI S., CAPECCHI B., MENEVERI R., PIRAS A.,  
RA LEUZZI R.;  
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN 12  
RP SEQUENCE FROM N.A.  
RA VATERONI L., MUSIO A., MENEVERI R., RAINALDI G.;  
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL: U74487; G2581764; -.  
DR PROSITE: PS00348; P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
FT NON\_TER 1  
FT 1  
SQ SEQUENCE 205 AA: 23122 MW; 680DDDDC CRC32;

Query Match 79.4%; Score 54; DB 10; Length 205;  
Best Local Similarity 80.0%; Pred. No. 1.05e-01;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 174 LPPGSKRAL 183  
1111111111  
QY 1 LPPGSKRAL 10

RESULT 15  
ID P89004 PRELIMINARY: PRT: 238 AA.  
AC P89004:  
DT 01-MAY-1997 (TREMBLERL. 03, CREATED)  
DT 01-MAY-1997 (TREMBLERL. 03, LAST SEQUENCE UPDATE)  
DT 01-MAY-1997 (TREMBLERL. 03, LAST ANNOTATION UPDATE)  
DE P53 (FRAGMENT).  
OS MASTOMYS NATALENSIS PAPILLOMAVIRUS (MNPV).  
OC VIRIDAE: DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.  
RN 11  
RP SEQUENCE FROM N.A.  
RC TISSUE-ECTOMA INDUCED BY LOXTIDINE.;  
RA LUDE E.A., TANG L.H., MODLIN I.M.;  
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U48618; G1813455; -.  
FT NON\_TER 1  
FT 1  
SQ SEQUENCE 238 AA: 26704 MW; 097E01F9 CRC32;

Query Match 77.9%; Score 53; DB 11; Length 238;  
Best Local Similarity 80.0%; Pred. No. 1.89e-01;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 192 LPIGSAKRAL 201  
1111111111  
QY 1 LPPGSKRAL 10

Search completed: Fri Sep 11 13:46:28 1998  
Job time : 39 secs.

**THIS PAGE BLANK (USPTO)**



# WARNING

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Search - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:40:05 1998; Maspar time 2.66 Seconds  
54.744 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-17  
Description: (1-9) from US08452843.pep  
Perfect Score: 64  
Sequence: 1 RPIILITL 9

Scoring table:  
PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: a-genes32

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 16.209; Variance 50.482; scale 0.321

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length DB | ID     | Description           | Pred. No. |
|------------|-------|-------------|-----------|--------|-----------------------|-----------|
| 1          | 64    | 100.0       | 113 10    | R51877 | Human p53 amino acids | 1.96e+00  |
| 2          | 64    | 100.0       | 157 10    | R51878 | Human p53 amino acids | 1.96e+00  |
| 3          | 64    | 100.0       | 319 24    | W28496 | Human p53 protein var | 1.96e+00  |
| 4          | 64    | 100.0       | 319 24    | W28495 | Human p53 protein var | 1.96e+00  |
| 5          | 64    | 100.0       | 335 24    | W28498 | Human p53 protein var | 1.96e+00  |
| 6          | 64    | 100.0       | 335 24    | W28497 | Human p53 protein var | 1.96e+00  |
| 7          | 64    | 100.0       | 337 21    | W13962 | Chimeric p53 protein. | 1.96e+00  |
| 8          | 64    | 100.0       | 333 24    | W28494 | Human p53 protein var | 1.96e+00  |
| 9          | 64    | 100.0       | 353 24    | W28493 | Human p53 protein var | 1.96e+00  |
| 10         | 64    | 100.0       | 361 21    | W13961 | Chimeric p53 protein. | 1.96e+00  |
| 11         | 64    | 100.0       | 361 21    | W13958 | Chimeric p53 protein. | 1.96e+00  |
| 12         | 64    | 100.0       | 363 24    | W28479 | Human p53 protein var | 1.96e+00  |
| 13         | 64    | 100.0       | 363 24    | W28480 | Human p53 protein var | 1.96e+00  |
| 14         | 64    | 100.0       | 363 21    | W13974 | Modified p53 variant  | 1.96e+00  |
| 15         | 64    | 100.0       | 363 21    | W13973 | Modified p53 variant  | 1.96e+00  |
| 16         | 64    | 100.0       | 363 21    | W13976 | Modified p53 variant  | 1.96e+00  |
| 17         | 64    | 100.0       | 370 21    | W13957 | Chimeric p53 protein. | 1.96e+00  |
| 18         | 64    | 100.0       | 374 24    | W28482 | Human p53 protein var | 1.96e+00  |

|    |    |       |        |        |                       |          |
|----|----|-------|--------|--------|-----------------------|----------|
| 19 | 64 | 100.0 | 374 24 | W28481 | Human p53 protein var | 1.96e+00 |
| 20 | 64 | 100.0 | 381 24 | W28489 | Human p53 protein var | 1.96e+00 |
| 21 | 64 | 100.0 | 381 24 | W28490 | Human p53 protein var | 1.96e+00 |
| 22 | 64 | 100.0 | 390 19 | W02623 | Mouse p53 protein.    | 1.96e+00 |
| 23 | 64 | 100.0 | 393 22 | W25155 | Human p53 variant fou | 1.96e+00 |
| 24 | 64 | 100.0 | 393 22 | W13951 | Human tumour-derived  | 1.96e+00 |
| 25 | 64 | 100.0 | 393 22 | W13948 | Human wild-type p53 t | 1.96e+00 |
| 26 | 64 | 100.0 | 393 21 | W05345 | Human p53 mutant N239 | 1.96e+00 |
| 27 | 64 | 100.0 | 393 21 | W05344 | Human p53 mutant R248 | 1.96e+00 |
| 28 | 64 | 100.0 | 393 21 | W05346 | Human p53 mutant R273 | 1.96e+00 |
| 29 | 64 | 100.0 | 393 21 | W05347 | Human p53 mutant R282 | 1.96e+00 |
| 30 | 64 | 100.0 | 393 21 | W05348 | Human p53 mutant R282 | 1.96e+00 |
| 31 | 64 | 100.0 | 393 18 | R91933 | Wild type p53 protein | 1.96e+00 |
| 32 | 64 | 100.0 | 393 19 | W02617 | Human p53 tumour supp | 1.96e+00 |
| 33 | 64 | 100.0 | 393 21 | W13968 | Modified p53 variant  | 1.96e+00 |
| 34 | 64 | 100.0 | 393 21 | W13970 | Modified p53 variant  | 1.96e+00 |
| 35 | 64 | 100.0 | 393 21 | W13969 | Modified p53 variant  | 1.96e+00 |
| 36 | 64 | 100.0 | 401 24 | W28487 | Human p53 protein var | 1.96e+00 |
| 37 | 64 | 100.0 | 401 24 | W28488 | Human p53 protein var | 1.96e+00 |
| 38 | 64 | 100.0 | 402 21 | W13965 | Chimeric p53 protein. | 1.96e+00 |
| 39 | 64 | 100.0 | 404 21 | W13963 | Chimeric p53 protein. | 1.96e+00 |
| 40 | 64 | 100.0 | 406 21 | W13966 | Chimeric p53 protein. | 1.96e+00 |
| 41 | 64 | 100.0 | 406 21 | W13964 | Chimeric p53 protein. | 1.96e+00 |
| 42 | 64 | 100.0 | 411 21 | W13967 | Chimeric p53 protein. | 1.96e+00 |
| 43 | 64 | 100.0 | 533 23 | W19763 | p53-GM-CSF immunostim | 1.96e+00 |
| 44 | 64 | 100.0 | 535 24 | W28492 | Human p53 protein var | 1.96e+00 |
| 45 | 64 | 100.0 | 535 24 | W28491 | Human p53 protein var | 1.96e+00 |

## ALIGNMENTS

RESULT 1  
AC R51877 standard; Protein: 113 AA.  
ID R51877;  
DE 18-NOV-1994 (first entry)  
KW Human p53 amino acids 237-349.  
OS Human nuclear phosphoprotein p53; tumour suppressor gene product;  
KW anti-oncogene; cancer; tumour; antibody binding region; epitope.  
KW Homo sapiens.  
FH Key Location/Qualifiers  
FT misc\_difference 37 /note= "Arg corresponds to a CAT codon"  
FT FT  
PD W09408241-A.  
PD 14-APR-1994.  
PF 30-SEP-1993; E02666.  
PR 30-SEP-1992; DE-232823.  
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
PI Klein R, Schanz P, Tessmer C, Volkmann M, Zentgraf H;  
DR WPI: 94-135732/16.  
DR N-PSDB: 062362.  
PT Non-radioactive detection of p53 specific antibodies - by capture  
PT on immobilised p53 or its fragments, then reaction with labelled  
PT second antibody, for diagnosis of tumours and suitable for  
PT screening  
PS Claim 10; Page 19; 35pp; German.  
CC Antibodies specific for p53 are detected by binding to immobilised  
CC fragments of the p53 gene product containing the antibody-binding  
CC region. Preferred fragments contain amino acids 1-241, 40-349,  
CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
CC 368-386. See R51872-R51881 for sequences of these fragments.  
SQ Sequence 113 AA;

Query Match 100.0%; Score 64; DB 10; Length 113;  
Best Local Similarity 100.0%; Pred. No. 1.96e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 13 RPIILITL 21  
QY 1 RPIILITL 9

RESULT 2  
ID R51878 standard; Protein: 157 AA.

AC R51878;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 237-393.  
 KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 KM anti-oncogene; cancer; tumour; antibody binding region; epitope.  
 OS Homo sapiens.  
 FH Key  
 FT misc\_difference 37  
 Location/Qualifiers  
 FT 1  
 /note- "Arg corresponds to a CAT codon"  
 PN W09408241-A.  
 PD 14-APR-1994.  
 PF 30-SEP-1993; E02666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schranz P, Tessmer C, Volkmann M, Zentgraf H;  
 DR N-PSDB; Q62363.  
 DR WPI: 94-135732/16.  
 PT Non-radioactive detection of p53 specific antibodies - by capture  
 PT on immobilised p53 or its fragments, then reaction with labelled  
 PT second antibody, for diagnosis of tumours and suitable for  
 PT screening  
 PS Claim 10; Page 19; 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 SQ Sequence 157 AA;  
 Query Match 100.0%; Score 64; DB 10; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 1.96e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 13 rplttltl 21  
 OY 1 RPLTTTL 9  
 RESULT 3  
 ID W28495;  
 AC W28495;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360-325H.  
 KW leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KM substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 OS Homo sapiens.  
 OS Synthetic.  
 FH Key  
 FT misc\_difference 145  
 Location/Qualifiers  
 FT 1  
 /note- "Arg residue at position 182 of wild-type  
 FT 1  
 p53 has been mutated to His"  
 PN W09704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON) RHONE-POULENC RORER SA.  
 PI Bracco L, Conseilier E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 38; Page -; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360-325H and comprising  
 CC the 325-360 domain, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors

CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 360-325).  
 SQ Sequence 319 AA;  
 Query Match 100.0%; Score 64; DB 24; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 1.96e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 213 rplttltl 221  
 OY 1 RPLTTTL 9  
 RESULT 4  
 ID W28495; standard; Protein: 319 AA.  
 AC W28495;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360-325 encoded by p5C178.  
 KW leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KM substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 OS Homo sapiens.  
 OS Synthetic.  
 PN W09704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON) RHONE-POULENC RORER SA.  
 PI Bracco L, Conseilier E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB; T86223.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 38; Pages 92-94; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360-325 and comprising  
 CC the 325-360 domain, amino acids 75-325 of human wild-type p53 and a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 319 AA;  
 Query Match 100.0%; Score 64; DB 24; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 1.96e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 213 rplttltl 221  
 OY 1 RPLTTTL 9  
 RESULT 5  
 ID W28498; standard; Protein: 335 AA.  
 AC W28498;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360h-325H.  
 KW leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KM substitution; replacement; transactivation; hinge region;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;

KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 FT Key Location/Qualifiers  
 FT region 39..53  
 FT /label= hinge  
 FT misc\_difference 161  
 FT /note= "Arg residue at position 182 of wild-type  
 p53 has been mutated to His"  
 PN MO9704092-A1.  
 PD 06-FEB-1997.  
 PE 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Concellier E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 leucine zipper - useful for treating hyper-proliferative disorders,  
 esp. cancer and restenosis  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360h-325h and comprising  
 CC the 325-360 domain, separated from amino acids 75-325 of human  
 CC wild-type p53 (but with Arg182 replaced by His) by a synthetic hinge  
 CC sequence (GlySer)<sub>3</sub>, and with a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 360h-325).  
 SO Sequence 335 AA;

Query Match 100.0%; Score 64; DB 24; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 1.96e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 229 RPLITITL 237  
 |||||||  
 Oy 1 RPLITITL 9

RESULT 6  
 ID W28497 standard; Protein; 335 AA.  
 AC W28497.  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360h-325 encoded by PEC179.  
 KW leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; hinge region;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 FT Key Location/Qualifiers  
 FT region 39..53  
 FT /label= hinge  
 PN MO9704092-A1.  
 PD 06-FEB-1997.  
 PE 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Concellier E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 leucine zipper - useful for treating hyper-proliferative disorders,  
 esp. cancer and restenosis

PS Claim 39: Pages 94-95; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360h-325h and comprising  
 CC the 325-360 domain, separated from amino acids 75-325 of human  
 CC wild-type p53 by a synthetic hinge sequence (GlySer)<sub>3</sub>, and with a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SO Sequence 335 AA;

Query Match 100.0%; Score 64; DB 24; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 1.96e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 229 RPLITITL 237  
 |||||||  
 Oy 1 RPLITITL 9

RESULT 7  
 ID W13962 standard; Protein; 337 AA.  
 AC W13962.  
 DT 25-JUN-1997 (first entry)  
 DE Chimeric p53 protein.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; GCN4; DNA binding.  
 OS Chimeric Homo sapiens;  
 OS Chimeric synthetic.  
 FT Key Location/Qualifiers  
 FT region 1..300  
 FT /label= p53wt  
 FT /note= "amino acids 1-300 of wild-type p53"  
 FT region 301..305  
 FT /label= linker  
 FT region 306..337  
 FT /label= GCN4  
 FT /note= "amino acids 250-281 of GCN4 LZ variant"  
 PN MO9710843-A1.  
 PD 27-MAR-1997.  
 PE 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI: 97-202618/18.  
 PT R284X modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Disclosure; Refer to Page 8; 82pp; English.  
 CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
 CC of human wild-type p53 tumour suppressor (see also W13955) linked  
 CC to a C-terminal portion of the LZ variant (see also W13955) of  
 CC GCN4 and, in some cases, the C-terminal portion of wild-type  
 CC p53. The chimeric proteins have DNA binding activity and can  
 CC replace lost or insufficient p53 function, providing the means for  
 CC pharmacological rescue of p53 function in cancer patients. Nucleic  
 CC acids coding for modified p53 constructs can be used for cancer  
 CC gene therapy.  
 SO Sequence 337 AA;

Query Match 100.0%; Score 64; DB 21; Length 337;  
 Best Local Similarity 100.0%; Pred. No. 1.96e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 249 RPLITITL 257  
 |||||||  
 Oy 1 RPLITITL 9

RESULT 8  
ID W28494 standard: Protein: 353 AA.  
AC W28494;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 393-325H.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muteln;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT misc-difference 179 /note="Arg residue at position 182 of wild-type  
FT p53 has been mutated to His"  
PN MO9704092-A1.  
PD 06-FEB-1997.  
PE 17-JUL-1996: F01111.  
PR 19-JUL-1995: FR-008729  
PA (RHON) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
PI WPI: 97-132633/12.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 37, Page -; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-393 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 393-325H and comprising  
CC the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with  
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
CC The p53 variants are more active and more stable tumour suppressors  
CC and apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not, i.e. they are not inactivated by dominant  
CC negative or oncogenic mutants, nor by other cellular proteins (because  
CC the leucine zipper domain prevents formation of inactive mixed  
CC oligomers).  
CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant 393-325).  
SQ Sequence 353 AA;

Query Match 100.0%; Score 64; DB 24; Length 353;  
Best Local Similarity 100.0%; Pred. No. 1.96e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 247 rp1ltltl 255  
| | | | | | | | | |  
OY 1 RP1LTTL 9

RESULT 9  
ID W28493 standard: Protein: 353 AA.  
AC W28493;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 393-325 encoded by p53177.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muteln;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
PN MO9704092-A1.  
PD 06-FEB-1997.  
PE 17-JUL-1996: F01111.  
PR 19-JUL-1995: FR-008729.  
PA (RHON) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
PI WPI: 97-132633/12.

DR N-PSDB: T86222.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 37, Pages 90-92; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-393 of p53. The present sequence is that of  
CC a specifically claimed p53 variant designated 393-325 and comprising  
CC the 325-393 domain, amino acids 75-325 of human wild-type p53 and a  
CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).  
SQ Sequence 353 AA;

Query Match 100.0%; Score 64; DB 24; Length 353;  
Best Local Similarity 100.0%; Pred. No. 1.96e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 247 rp1ltltl 255  
| | | | | | | | | |  
OY 1 RP1LTTL 9

RESULT 10  
ID W13961 standard: Protein: 361 AA.  
AC W13961;  
DT 25-JUN-1997 (first entry)  
DE Chimeric p53 protein.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; GCN4; DNA binding.  
OS Chimeric Homo sapiens.  
OS Chimeric synthetic.  
FH Key Location/Qualifiers  
FT region 1..323 /label="p53wt"  
FT region /note="amino acids 1-323 of wild-type p53"  
FT region 324..329 /label="Linker"  
FT region 330..361 /label="GCN4"  
FT /note="amino acids 250-281 of GCN4 LZ variant"  
PN MO9710843-A1.  
PD 27-MAR-1997.  
PE 20-SEP-1996: U15188.  
PR 22-SEP-1995: US-004802.  
PR 21-AUG-1996: US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
PI WPI: 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Disclosure: Refer to Page 8; 82pp; English.  
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
CC of human wild-type p53 tumour suppressor (see also W13948) linked  
CC to a C-terminal portion of the LZ variant (see also W13955) of  
CC GCN4 and, in some cases, the C-terminal portion of wild-type  
CC p53. The chimeric proteins have DNA binding activity and can  
CC replace lost or insufficient p53 function, providing the means for  
CC pharmacological rescue of p53 function in cancer patients. Nucleic  
CC acids coding for modified p53 constructs can be used for cancer  
CC gene therapy.  
SQ Sequence 361 AA;

Query Match 100.0%; Score 64; DB 21; Length 361;  
Best Local Similarity 100.0%; Pred. No. 1.96e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Sun Sep 13 10:55:42 1998

```

DB 249 RpLITLTL 257
QY 1 RpLITLTL 9

RESULT 11
ID W13958 standard; Protein: 361 AA.
AC W13958;
DE 25-JUN-1997 (first entry)
DE Chimeric p53 protein.
KW p53; tumour suppressor; cancer; therapy; cell proliferation;
KW apoptosis; protein engineering; GCN4; DNA binding.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT region 1..325
FT /label= p53wt
FT /note= "amino acids 1-325 of wild-type p53"
FT 326..328
FT /label= Linker
FT region 329..361
FT /label= GCN4
FT /note= "amino acids 249-281 of GCN4 LZ variant"
PN W09710843-A1.
PD 27-MAR-1997.
PF 20-SEP-1996; US15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazoneclis TD;
DR WPI: 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PS Disclosure; Refer to Page 8: 82pp; English.
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions
CC of human wild-type p53 tumour suppressor (see also W13948) linked
CC to a C-terminal portion of the LZ variant (see also W13955) of
CC GCN4 and, in some cases, the C-terminal portion of wild-type
CC p53. The chimeric proteins have DNA binding activity and can
CC replace lost or insufficient p53 function, providing the means for
CC pharmacological rescue of p53 function in cancer patients. Nucleic
CC acids coding for modified p53 constructs can be used for cancer
CC gene therapy.
SO Sequence 361 AA.

Query Match 100.0%; Score 64; DB 21; Length 361;
Best Local Similarity 100.0%; Pred. No. 1.96e+00;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 249 RpLITLTL 257
QY 1 RpLITLTL 9

RESULT 12
ID W28479 standard; Protein: 363 AA.
AC W28479;
DE 25-NOV-1997 (first entry)
DE Human p53 protein variant V-325 encoded by pEC114.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Herpes simplex virus.
OS Synthetic.
PN W09704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PF 19-JUL-1995; FR-008729.
PR 19-JUL-1995; FR-008729.
PA (RHON) RHONE-POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI: 97-132633/12.

```

```

DR N-PSDB; T86215.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 30; Pages 76-78; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the transactivating domain (TD) from herpes simplex virus viral
CC protein VP16 (amino acids 411-490). The present sequence is that of
CC a specifically claimed p53 variant designated V-325 and comprising
CC the VP16 TD, amino acids 75-325 of human wild-type p53 and a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
SO Sequence 363 AA.

Query Match 100.0%; Score 64; DB 24; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.96e+00;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 257 RpLITLTL 265
QY 1 RpLITLTL 9

RESULT 13
ID W28480 standard; Protein: 363 AA.
AC W28480;
DE 25-NOV-1997 (first entry)
DE Human p53 protein variant V-325H.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Herpes simplex virus.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_difference 189
FT /note= "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
FT PN W09704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PF 19-JUL-1995; FR-008729.
PR 19-JUL-1995; FR-008729.
PA (RHON) RHONE-POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI: 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 30; Page -; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the transactivating domain (TD) from herpes simplex virus viral
CC protein VP16 (amino acids 411-490). The present sequence is that of
CC a specifically claimed p53 variant designated V-325H and comprising
CC the VP16 TD, amino acids 75-325 of human wild-type p53 (but with
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.
CC The p53 variants are more active and more stable tumour suppressors
CC and apoptosis-inducing agents than wild-type p53 and are active where
CC the wild-type protein is not, i.e. they are not inactivated by dominant
CC negative or oncogenic mutants, nor by other cellular proteins (because
CC the leucine zipper domain prevents formation of inactive mixed
CC oligomers).
CC (Note: this sequence does not appear in the specification and has

```

CC been produced by modifying the given sequence of variant V-325).

SC Sequence 363 AA;

Query Match 100.0%; Score 64; DB 24; Length 363;

Best Local Similarity 100.0%; Pred. No. 1.96e+00; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 257 rplltlcl 265  
OY 1 RPLLTITL 9

RESULT 14

ID W13974 standard; Protein: 363 AA.

AC W13974;

DT 25-JUN-1997 (first entry)

DE Modified p53 variant p53H273del1364-393.

KW p53; tumour suppressor; cancer; therapy; cell proliferation;

KW apoptosis; protein engineering; DNA binding.

OS Synthetic.

PN W09710843-A1.

PD 27-MAR-1997

PF 20-SEP-1996; U15188.

PR 22-SEP-1995; US-004802.

PR 21-AUG-1996; US-697221.

PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.

PI Halazoneis TD;

DR WPI: 97-202618/18.

PT R284K modified p53 protein having DNA binding ability - useful in

treatment of cancer

PS Example 1; 56-57; 82pp; English.

CC Modified p53 variant p53H273del1364-393 (W13974) has the tumour-

derived histidine 273 mutation (see also W13952) and a deletion

of the C-terminal 30 amino acids of wild-type p53 (see also

W13948). His273 is a Class I p53 tumour mutation that affects DNA

binding. The C-terminal deletion, introduced by site-directed

mutagenesis of p53 DNA, activates the DNA binding of the p53

tumour mutant. This provides the means for pharmacological rescue

of p53 function in cancer patients. Other modified p53 constructs

(W13949-50, W13953-54, W13968-77) have also been produced. Nucleic

acids coding for modified p53 can be used for cancer gene therapy.

Sequence 363 AA;

Query Match 100.0%; Score 64; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 1.96e+00; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 249 rplltlcl 257  
OY 1 RPLLTITL 9

RESULT 15

ID W13973 standard; Protein: 363 AA.

AC W13973;

DT 25-JUN-1997 (first entry)

DE Modified p53 variant p53Q248R284del1364-393.

KW p53; tumour suppressor; cancer; therapy; cell proliferation;

KW apoptosis; protein engineering; DNA binding.

OS Synthetic.

PN W09710843-A1.

PD 27-MAR-1997

PF 20-SEP-1996; U15188.

PR 22-SEP-1995; US-004802.

PR 21-AUG-1996; US-697221.

PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.

PI Halazoneis TD;

DR R284K modified p53 protein having DNA binding ability - useful in

treatment of cancer

PS Example 1; 54-56; 82pp; English.

CC Modified p53 variant p53Q248R284del1364-393 (W13973) has the tumour-

derived Gln248 mutation (see also W13951), a Thr284 to Arg substn.

CC (see also W13949) and a deletion of the 30 C-terminal amino acids

CC of wild-type p53 (W13948). Gln248 is a Class I p53 tumour mutation

CC that affects DNA binding. The T284R substitution, introduced by

CC site-directed mutagenesis of p53 DNA, provides a novel p53-DNA

CC contact between a phosphate of the DNA backbone and p53, and

CC restores DNA binding. The C-terminal deletion permits in vitro

CC DNA binding. The construct provides the means for pharmacological

CC rescue of p53 function in cancer patients. Other modified p53

CC constructs (W13949-50, W13953-54, W13968-77) have also been

CC produced. Nucleic acids coding for modified p53 can be used for

CC cancer gene therapy.

Sequence 363 AA;

Query Match 100.0%; Score 64; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 1.96e+00; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 249 rplltlcl 257  
OY 1 RPLLTITL 9

Search completed: Fri Sep 11 13:40:15 1998

Job time : 10 secs.

\*\*\*\*\*  
WIREIMAGE  
(TM)  
\*\*\*\*\*

Release: JA John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:40:34 1998; MasPar time 3.56 Seconds

Tabular output not generated. 92.411 Million cell updates/sec

Title: >US-08-452-843-17  
Description: (1-9) from US08452843.pep  
Perfect Score: 64  
Sequence: 1 RPILITITL 9

Scoring table: PAM 150  
Gap 15

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r56  
1:p1r1 2:p1r2 3:p1r3 4:p1r4 5:nr13d

Statistics: Mean 23.053; Variance 31.139; scale 0.740

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description                           | Pred. No. |
|------------|-------|-------------|--------|----|---------------------------------------|-----------|
| 1          | 64    | 100.0       | 191    | 5  | 1YCSA p53 residues 97-287, 1.57e-02   |           |
| 2          | 64    | 100.0       | 194    | 5  | 1TUPB tumor suppressor p53, 1.57e-02  |           |
| 3          | 64    | 100.0       | 194    | 5  | 1TSCR tumor suppressor, 1.57e-02      |           |
| 4          | 64    | 100.0       | 195    | 5  | 1TUPC tumor suppressor p53, 1.57e-02  |           |
| 5          | 64    | 100.0       | 195    | 5  | 1TSCR tumor suppressor, 1.57e-02      |           |
| 6          | 64    | 100.0       | 196    | 5  | 1TUPA tumor suppressor p53, 1.57e-02  |           |
| 7          | 64    | 100.0       | 196    | 5  | 1TSCR tumor suppressor, 1.57e-02      |           |
| 8          | 64    | 100.0       | 196    | 5  | 1TSCR tumor suppressor, 1.57e-02      |           |
| 9          | 64    | 100.0       | 381    | 2  | S38824 cellular tumor antige 1.57e-02 |           |
| 10         | 64    | 100.0       | 386    | 2  | S51648 cellular tumor antige 1.57e-02 |           |
| 11         | 64    | 100.0       | 390    | 1  | DNMS53 cellular tumor antige 1.57e-02 |           |
| 12         | 64    | 100.0       | 391    | 2  | S02192 cellular tumor antige 1.57e-02 |           |
| 13         | 64    | 100.0       | 391    | 2  | JC6193 tumor suppressor p53 1.57e-02  |           |
| 14         | 64    | 100.0       | 393    | 1  | JC6176 tumor suppressor p53 1.57e-02  |           |
| 15         | 64    | 100.0       | 393    | 2  | DNH53 cellular tumor antige 1.57e-02  |           |
| 16         | 64    | 100.0       | 393    | 2  | S06594 cellular tumor antige 1.57e-02 |           |
| 17         | 64    | 100.0       | 396    | 2  | JH0633 cellular tumor antige 1.57e-02 |           |
| 18         | 64    | 100.0       | 396    | 2  | JH0631 cellular tumor antige 1.57e-02 |           |
| 19         | 59    | 92.2        | 367    | 2  | S02193 cellular tumor antige 1.79e-01 |           |
| 20         | 50    | 78.1        | 328    | 2  | A55215 Kdgr 5'-region hypoch 1.08e+01 |           |
| 21         | 50    | 78.1        | 522    | 2  | JC4532 cytochrome P450 4F4 p 1.08e+01 |           |
| 22         | 49    | 76.6        | 568    | 2  | JQ2206 Ula6h protein - Marek 1.67e+01 |           |
| 23         | 48    | 75.0        | 160    | 2  | I44020 Trbh - plasmid RK2 2.55e+01    |           |

|    |    |      |     |   |  |
|----|----|------|-----|---|--|
| 24 | 48 | 75.0 | 253 | 2 | B69758 conserved hypothetical 2.55e+01 |
| 25 | 48 | 75.0 | 397 | 2 | G70078 pyrimidine nucleoside 2.55e+01  |
| 26 | 48 | 75.0 | 533 | 2 | S71617 dimethylamine mono 2.55e+01     |
| 27 | 48 | 75.0 | 533 | 2 | S51131 flavin-containing mon 2.55e+01  |
| 28 | 48 | 75.0 | 533 | 2 | S71618 dimethylamine mono 2.55e+01     |
| 29 | 47 | 73.4 | 391 | 2 | H69213 conserved hypothetical 3.86e+01 |
| 30 | 47 | 73.4 | 459 | 2 | H64667 conserved hypothetical 3.86e+01 |
| 31 | 47 | 73.4 | 464 | 2 | S34024 alpha-1,2-mannosyltra 3.86e+01  |
| 32 | 46 | 71.9 | 100 | 2 | B36796 hypothetical protein 5.83e+01   |
| 33 | 46 | 71.9 | 333 | 2 | D69812 ferrichrome ABC trans 5.83e+01  |
| 34 | 46 | 71.9 | 439 | 2 | A64769 branched chain amino 5.83e+01   |
| 35 | 46 | 71.9 | 439 | 2 | J00007 brno protein - Salmon 5.83e+01  |
| 36 | 46 | 71.9 | 698 | 2 | S52674 general sporulation p 5.83e+01  |
| 37 | 46 | 71.9 | 730 | 2 | JH0798 fasciclin IV precursor 8.73e+01 |
| 38 | 45 | 70.3 | 157 | 2 | S38031 probable olfactory re 8.73e+01  |
| 39 | 45 | 70.3 | 222 | 2 | C40745 odorant receptor (clo 8.73e+01  |
| 40 | 45 | 70.3 | 226 | 1 | XJBY5 orotate phosphoribosy 8.73e+01   |
| 41 | 45 | 70.3 | 227 | 1 | XJBY10 orotate phosphoribosy 8.73e+01  |
| 42 | 45 | 70.3 | 321 | 2 | E69687 cytochrome aa3 quinol 8.73e+01  |
| 43 | 45 | 70.3 | 452 | 2 | A64933 cglB protein - Escher 8.73e+01  |
| 44 | 45 | 70.3 | 604 | 2 | F69802 ABC transporter (ATP- 8.73e+01  |
| 45 | 45 | 70.3 | 868 | 1 | VBEB31 glycoprotein B - huma 8.73e+01  |

## ALIGNMENTS

|                       |   |                |
|-----------------------|---|----------------|
| RESULT 1              | 1YCSA   | #type complete |
| ENTRY                 | p53 residues 97-287, chain A - human  |                |
| TITLE                 | p53-53bp2 complex   |                |
| PDB TITLE             | #formal name Homo sapiens #common name man  |                |
| ORGANISM              | expressed in Escherichia coli, strain b121 (d3)   |                |
| #note                 | A68208  |                |
| REFERENCE             | Gorina, S.; Pavletich, N.P.   |                |
| #authors              | submitted to the Brookhaven Protein Data Bank, September 1996                                     |                |
| #submission           | references PDB:1YCS   |                |
| REFERENCE             | TN001216  |                |
| #authors              | Gorina, S.; Pavletich, N.P.   |                |
| #journal              | Science (1996) 274:1001   |                |
| #title                | Structure of the p53 tumor suppressor bound to the ankyrin and sh3 domains of 53bp2.              |                |
| REFERENCE             | TN001217  |                |
| #authors              | Naumovski, L.; Cleary, M.L.   |                |
| #journal              | Mol. Cell. Biol. (1996) 16:3884   |                |
| #title                | The p53-binding protein 53bp2 also interacts with bcl2 and impedes cell cycle progression at g2m. |                |
| REFERENCE             | I38604  |                |
| #authors              | Iwabuchi, K.; Bartel, P.L.; Li, B.; Marracchino, R.; Fields, S.                                   |                |
| #journal              | Proc. Natl. Acad. Sci. U.S.A. (1994) 91:6098-6102   |                |
| #title                | Two cellular proteins that bind to wild-type but not mutant p53.                                  |                |
| COMMENT               | #cross-references MUID:94286584   |                |
| COMMENT               | Resolution: 2.2 angstroms   |                |
| COMMENT               | Determination: X-ray diffraction  |                |
| COMMENT               | R-value: 0.205  |                |
| KEYWORDS              | ankyrin repeats; anti-oncogene; complex; disease mutation   |                |
| FEATURE               | polymorphism; multigene family; nuclear protein; p53; phosphorylation; sh3; tumor suppressor      |                |
| 9-11                  | #region helix (right hand 3-10)\  |                |
| 81-85                 | #region helix (right hand alpha)\   |                |
| 182-190               | #region helix (right hand alpha)\   |                |
| 14-16,45-50,          | #region beta sheet\   |                |
| 134-140,99-101        |   |                |
| 28-31,36-39,          |   |                |
| 168-178,155-162,      |   |                |
| 60-67,118-123,        |   |                |
| 108-111               |   |                |
| SUMMARY               | #length 191 #molecular-weight 21515 #checksum 8219  |                |
| Query Match           | 100.0%; Score 64; DB 5; Length 191;   |                |
| Best Local Similarity | 100.0%; Pred. No. 1.57e-02;   |                |

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 153 RPLTITL 161  
 QY 1 RPLTITL 9

RESULT 2

ENTRY 1TUPB #type complete  
 TITLE tumor suppressor p53, chain B - human  
 PDB-TITLE tumor suppressor p53 complexed with DNA  
 ORGANISM #formal\_name Homo sapiens #common\_name man  
 #note expressed in Escherichia coli  
 A66776

REFERENCE  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
 #submission submitted to the Brookhaven Protein Data Bank, July 1995  
 #cross-references PDB:1TUP  
 A43072

REFERENCE  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
 #journal Science (1994) 265:346-355  
 #title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.  
 A49450

REFERENCE  
 #authors Pavletich, N.P.; Chambers, K.A.; Pabo, C.O.  
 #journal Genes Dev. (1993) 7:2556-2564  
 #title The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots.  
 TN031795

REFERENCE  
 #authors Vogelstein, B.; Kinzler, K.W.  
 #journal Cell (1992) 70:523  
 #title p53 function and dysfunction.  
 #note

COMMENT Resolution: 2.2 angstroms  
 DETERMINATION: X-ray diffraction  
 R-value: 0.202

KEYWORDS antigen p53; complex; DNA; tumor suppressor

FEATURE  
 72-75 #region helix (right hand 3-10)\  
 82-86 #region helix (right hand alpha)\  
 153-192 #region helix (right hand alpha)\  
 15-17,46-51, #region helix (right hand alpha)\  
 135-141,100-102 #region beta sheet\  
 29-32,37-40,  
 169-179,156-163,  
 61-68,119-124,  
 109-112

SUMMARY #region beta sheet  
 #length 194 #molecular-weight 21830 #checksum 2852

Query Match 100.0%; Score 64; DB 5; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 1.57e-02;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 RPLTITL 162  
 QY 1 RPLTITL 9

RESULT 3

ENTRY 1TUPB #type complete  
 TITLE p53 tumor suppressor, chain B - human  
 PDB-TITLE p53 core domain in complex with DNA  
 ORGANISM #formal\_name Homo sapiens #common\_name man  
 #note expressed in Escherichia coli  
 A66760

REFERENCE  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.; Pavletich, N.  
 #submission submitted to the Brookhaven Protein Data Bank, July 1995  
 #cross-references PDB:1TUP  
 A43072

REFERENCE  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
 #journal Science (1994) 265:346-355  
 #title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.  
 Resolution: 2.2 angstroms  
 DETERMINATION: X-ray diffraction

KEYWORDS anti-oncogene; complex; DNA; DNA-binding protein; tumor suppressor

FEATURE  
 72-75 #region helix (right hand 3-10)\  
 82-86 #region helix (right hand alpha)\  
 183-192 #region helix (right hand alpha)\  
 15-17,46-51, #region beta sheet\  
 135-141,100-103  
 29-32,37-40,  
 169-179,156-163,  
 61-68,119-124,  
 109-112

SUMMARY #region beta sheet  
 #length 194 #molecular-weight 21830 #checksum 2852

Query Match 100.0%; Score 64; DB 5; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 1.57e-02;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 RPLTITL 162  
 QY 1 RPLTITL 9

RESULT 4

ENTRY 1TUPC #type complete  
 TITLE tumor suppressor p53, chain C - human  
 PDB-TITLE tumor suppressor p53 complexed with DNA  
 ORGANISM #formal\_name Homo sapiens #common\_name man  
 #note expressed in Escherichia coli  
 A66776

REFERENCE  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
 #submission submitted to the Brookhaven Protein Data Bank, July 1995  
 #cross-references PDB:1TUP  
 A43072

REFERENCE  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
 #journal Science (1994) 265:346-355  
 #title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.  
 A49450

REFERENCE  
 #authors Pavletich, N.P.; Chambers, K.A.; Pabo, C.O.  
 #journal Genes Dev. (1993) 7:2556-2564  
 #title The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots.  
 TN031798

REFERENCE  
 #authors Vogelstein, B.; Kinzler, K.W.  
 #journal Cell (1992) 70:523  
 #title p53 function and dysfunction.  
 #note

COMMENT Resolution: 2.2 angstroms  
 DETERMINATION: X-ray diffraction  
 R-value: 0.202

KEYWORDS antigen p53; complex; DNA; tumor suppressor

FEATURE  
 11-13 #region helix (right hand 3-10)\  
 72-75 #region helix (right hand alpha)\  
 83-87 #region helix (right hand alpha)\  
 184-191 #region helix (right hand alpha)\  
 16-19,47-52,  
 136-142,101-103 #region beta sheet\  
 31-33,38-42,  
 170-181,157-164,  
 62-69,120-125,  
 110-113

SUMMARY #region beta sheet  
 #length 195 #molecular-weight 21917 #checksum 4657

Query Match 100.0%; Score 64; DB 5; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 1.57e-02;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 155 RPLTITL 163  
 QY 1 RPLTITL 9



RESULT 5  
ENTRY 1TSRC #type complete  
TITLE p53 tumor suppressor, chain C - human  
PDB\_TITLE p53 core domain in complex with DNA  
ORGANISM #formal\_name Homo sapiens #common\_name man  
#note expressed in Escherichia coli  
REFERENCE A66760  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.; Pavletich, N.  
#submission submitted to the Brookhaven Protein Data Bank, July 1995  
REFERENCE A43072  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex:  
understanding tumorigenic mutations.  
COMMENT Resolution: 2.2 angstroms  
Determination: X-ray diffraction  
KEYWORDS anti-oncogene; complex; DNA; DNA-binding protein; tumor  
suppressor  
FEATURE  
11-13 #region helix (right hand 3-10) \  
72-75 #region helix (right hand alpha) \  
83-87 #region helix (right hand alpha) \  
184-191 #region helix (right hand alpha) \  
16-19,47-52, #region beta sheet \  
136-142,101-103  
31-33,38-42,  
170-181,157-164,  
62-69,120-125,  
110-113  
SUMMARY #length 195 #molecular-weight 21917 #checksum 4657  
Query Match 100.0%; Score 64; DB 5; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1.57e-02;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 155 RPITITL 163  
QY 1 RPITITL 9  
RESULT 6  
ENTRY 1TUPA #type complete  
TITLE tumor suppressor p53, chain A - human  
PDB\_TITLE tumor suppressor p53 complexed with DNA  
ORGANISM #formal\_name Homo sapiens #common\_name man  
#note expressed in Escherichia coli  
REFERENCE A66776  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#submission submitted to the Brookhaven Protein Data Bank, July 1995  
REFERENCE A43072  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex:  
understanding tumorigenic mutations.  
REFERENCE A49450  
#authors Pavletich, N.P.; Chambers, K.A.; Pabo, C.O.  
#journal Genes Dev. (1993) 7:2556-2564  
#title The DNA-binding domain of p53 contains the four conserved  
regions and the major mutation hot spots.  
REFERENCE TNO31792  
#authors Vogelstein, B.; Kinzler, K.W.  
#journal Cell (1992) 70:523  
#title p53 function and dysfunction.  
COMMENT Resolution: 2.2 angstroms  
Determination: X-ray diffraction  
KEYWORDS R-value: 0.202  
antigen p53; complex; DNA; tumor suppressor  
FEATURE  
73-75 #region helix (right hand 3-10) \  
84-87 #region helix (right hand alpha) \  
185-194 #region helix (right hand alpha) \  
Db 155 RPITITL 164  
QY 1 RPITITL 9

17-19,48-53, #region beta sheet \  
137-143,102-105  
31-34,39-42,  
171-181,158-165,  
63-70,121-126,  
111-114  
SUMMARY #length 196 #molecular-weight 22004 #checksum 7058  
Query Match 100.0%; Score 64; DB 5; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.57e-02;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 156 RPITITL 164  
QY 1 RPITITL 9  
RESULT 7  
ENTRY 1TSRA #type complete  
TITLE p53 tumor suppressor, chain A - human  
PDB\_TITLE p53 core domain in complex with DNA  
ORGANISM #formal\_name Homo sapiens #common\_name man  
#note expressed in Escherichia coli  
REFERENCE A66760  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.; Pavletich, N.  
#submission submitted to the Brookhaven Protein Data Bank, July 1995  
REFERENCE A43072  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex:  
understanding tumorigenic mutations.  
COMMENT Resolution: 2.2 angstroms  
Determination: X-ray diffraction  
KEYWORDS anti-oncogene; complex; DNA; DNA-binding protein; tumor  
suppressor  
FEATURE  
73-75 #region helix (right hand 3-10) \  
84-87 #region helix (right hand alpha) \  
185-194 #region helix (right hand alpha) \  
17-19,48-53, #region beta sheet \  
137-143,102-105  
31-34,39-42,  
171-181,158-165,  
63-70,121-126,  
111-114  
SUMMARY #length 196 #molecular-weight 22004 #checksum 7058  
Query Match 100.0%; Score 64; DB 5; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.57e-02;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 156 RPITITL 164  
QY 1 RPITITL 9  
RESULT 8  
ENTRY A29376 #type complete  
TITLE cellular tumor antigen p53 - African clawed frog  
ORGANISM #formal\_name Xenopus laevis #common\_name African clawed frog  
DATE 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change  
08-Sep-1997  
ACCESSIONS A29376; S61531; S72313; I51639  
REFERENCE A29376  
#authors Soussi, T.; de Fromental, C.C.; Mechali, M.; May, P.; Kress,  
M.  
#journal Oncogene (1987) 1:71-78  
#title Cloning and characterization of a cDNA from Xenopus laevis  
coding for a protein homologous to human and murine p53.  
#cross-references MIM:88143684  
#accession A29376  
#molecule\_type mRNA

```

#residues 1-363 ##label SOU
##cross-references EMBL:X05191; NID:g64961; PID:g64962
REFERENCE 151639
#authors Hoeber, M.; Clement, J.H.; Wedlich, D.; Montenarh, M.;
#journal Oncogene (1994) 9:109-120
#title Overexpression of wild-type p53 interferes with normal
#cross-references MUID:94134403
#accession S61531
##molecule-type mRNA
##residues 1-293,295-363 ##label HOE
##cross-references EMBL:X77546; NID:g468513; PID:g468514
REFERENCE S72313
#authors Hoeber, M.; Clement, J.; Wedlich, D.; Montenarh, M.; Knochel,
#submission submitted to the EMBL Data Library, March 1994
#accession S72313
##molecule-type mRNA
##residues 1-51,'S',53-70,72-293,295-363 ##label HOW
##cross-references EMBL:X77546; NID:g468513; PID:g468514
GENETICS
#gene p53
#CLASSIFICATION superfamily cellular tumor antigen p53
#KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
FEATURE
150,153,213,217 #binding-site zinc (Cys, His, Cys, Cys) #status
predicted\
362 #binding-site phosphoryl-RNA (Ser) (covalent) #status
predicted
SUMMARY #length 363 #molecular-weight 40692 #checksum 6648
Query Match 100.0%; Score 64; DB 2; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.57e-02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 RPLITITL 9
Db 224 RPLITITL 232
Oy 1 RPLITITL 9
RESULT 9
ENTRY S38824 #type complete
TITLE cellular tumor antigen p53, alternative splice form - mouse
#formal_name Mus musculus #common_name mouse
#ORGANISM 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
25-Oct-1996
#accessions S38824; S35478
REFERENCE S38822
#authors Arat, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:3332-3239
#title Immunologically distinct p53 molecules generated by
alternative splicing.
#accession S38824
##molecule-type mRNA
##residues 1-381 ##label ARA
REFERENCE S35478
#authors Han, K.A.; Kulesz-Martin, M.F.
#journal Nucleic Acids Res. (1992) 20:1979-1981
#title Alternatively spliced p53 RNA in transformed and normal cells
of different tissue types.
#accession S35478
#status nucleic acid sequence not shown; translation not shown
#molecule-type mRNA
#residues 1-381 ##label HAN
#cross-references EMBL:M3874; NID:g200202; PID:g200203
#note The nucleotide sequence was submitted to the EMBL Data
Library, July 1988
COMMENT This sequence, produced by alternative splicing of the tenth
intron, lacks the carboxyl-terminal sequence necessary for

```

```

covalent attachment of RNA. The function of this minor splice
form is not known.
CLASSIFICATION #superfamily cellular tumor antigen p53
#KEYWORDS alternative splicing; phosphoprotein
FEATURE
1-44
#domain transcription activation #status predicted
#label TRA\
16-26 #region conserved region I\
99-289 #domain DNA-binding core #status predicted #label DBC\
108-121 #region L1 loop\
114-139 #region conserved region II\
160-192 #region conserved region IIV\
168-178 #region conserved region IIV\
211-252 #region conserved region IV\
233-248 #region L3 loop\
267-283 #region conserved region V\
313-319 #region nuclear location signal\
319-357 #region tetramer association\
7,9,12,18,23,37 #binding-site phosphate (Ser) (covalent) #status
predicted\
173,176,235,239 #binding-site zinc (Cys, His, Cys, Cys) #status
predicted\
312 #binding-site phosphate (Ser) (covalent) (by cdc2
kinase) #status predicted
SUMMARY #length 381 #molecular-weight 42498 #checksum 8703
Query Match 100.0%; Score 64; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.57e-02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 RPLITITL 9
Db 246 RPLITITL 254
Oy 1 RPLITITL 9
RESULT 10
ENTRY S51648 #type complete
TITLE tumor-suppressor protein p53 - bovine
#formal_name Bos primigenius taurus #common_name cattle
#ORGANISM 07-May-1995 #sequence_revision 01-Sep-1995 #text_change
08-Sep-1997
#accessions S51648
REFERENCE S51648
#authors Deglert, F.; Willems, L.; Burny, A.; Kettmann, R.
#submission submitted to the EMBL Data Library, September 1994
#description Nucleotide sequence of the ovine p53 tumor-suppressor gene
cDNA and its genomic organisation.
#accession S51648
##molecule-type mRNA
##residues 1-386 ##label DEQ
##cross-references EMBL:X81704; NID:g602332; PID:g602333
CLASSIFICATION #superfamily cellular tumor antigen p53
#KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
phosphoprotein; transcription regulation; tumor suppressor;
zinc
FEATURE
168,171,231,235 #binding-site zinc (Cys, His, Cys, Cys) #status
predicted\
385 #binding-site phosphoryl-RNA (Ser) (covalent) #status
predicted
SUMMARY #length 386 #molecular-weight 43255 #checksum 7025
Query Match 100.0%; Score 64; DB 2; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.57e-02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 RPLITITL 9
Db 242 RPLITITL 250
Oy 1 RPLITITL 9

```

```

RESULT 11
ENTRY #type complete
TITLE cellular tumor antigen p53 - mouse
ALTERNATE_NAMES oncoprotein p53
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Aug-1985 #sequence_revision 04-Oct-1996 #text_change
05-Sep-1997
ACCESSIONS A22739; S06336; A02684; S38822; S38823; I48703
REFERENCE A22739
#authors Bienz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.
#journal EMBO J. (1984) 3:2179-2183
#cross-references M01D:85027173
#accession A22739
#molecule_type DNA
#residues 1-134, 'V', 136-390 ##label BIE
REFERENCE S06336
#authors Chumakov, P.M.
#journal Bioorg. Khim. (1987) 13:1691-1694
#title Primary structure of DNA complementary to murine oncoprotein
#accession S06336
#cross-references M01D:88221682
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-134, 'V', 136-390 ##label CHU
REFERENCE A02684
#authors Zakut-Houri, R.; Oren, M.; Bienz, B.; Lavie, V.; Hazum, S.;
Givol, D.
#journal Nature (1983) 306:594-597
#title A single gene and a pseudogene for the cellular tumor
antigen p53.
#cross-references M01D:84068204
#accession A02684
#molecule_type mRNA
#residues 1-159, 'H', 161-167, 'G', 169-233, 'T', 235-390 ##label ZAK
REFERENCE S38822
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:3232-3239
#title Immunologically distinct p53 molecules generated by
alternative splicing.
#accession S38822
#status preliminary
#molecule_type mRNA
#residues 1-390 ##label ARA
#cross-references EMBL:M1872; M1D:9200198; PID:9200199
#accession S38823
#status preliminary
#molecule_type mRNA
#residues 1-167, 'G', 169-233, 'T', 235-390 ##label AR2
#cross-references EMBL:M1873
REFERENCE I48703
#authors Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.
#journal Nucleic Acids Res. (1984) 12:5609-5626
#title Cloning and expression analysis of full length mouse cDNA
sequences encoding the transformation associated protein
p53.
#cross-references M01D:84272240
#accession I48703
#status preliminary; translated from GB/EMBL/DBD
#molecule_type mRNA
#residues 1-47, 'R', 49-78, 'QW', 82-390 ##label RRS
#cross-references EMBL:X00741; M1D:953570; PID:953571
COMMENT This DNA-binding protein plays an essential role in the regulation
of cell division, as it is required for the transition from phase
G0 to G1 of the cell cycle.
The tetramer association region may exhibit a beta-turn,
beta-sheet, beta-turn, alpha-helix motif.
COMMENT
CLASSIFICATION
#superfamily cellular tumor antigen p53
#apoptosis; cell division control; DNA binding; homotetramer;
phosphoprotein; transcription regulation; tumor suppressor;
zinc
FEATURE

```

```

1-44
16-26 #domain transcription activation #status predicted
99-289 #region conserved region I\
108-121 #domain DNA-binding core #status predicted #label DBC\
114-139 #region L1 loop\
160-192 #region conserved region II\
168-178 #region L2 loop\
231-252 #region conserved region III\
233-248 #region conserved region IV\
267-283 #region L3 loop\
313-319 #region conserved region V\
319-357 #region nuclear location signal\
7,9,12,18,23,37 #region tetramer association\
#binding_site phosphate (Ser) (covalent) #status
predicted\
173,176,235,239 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
312 #binding_site phosphate (Ser) (covalent) (by cdcc
kinase) #status predicted\
389 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted\
SUMMARY #length 390 #molecular_weight 43458 #checksum 1260
Query Match 100.0%; Score 64; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 1,57e-02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 246 RPIITITL 254
QY 1 RPIITITL 9
RESULT 12
ENTRY S02192
TITLE cellular tumor antigen p53 - rat
ALTERNATE_NAMES gene p53 protein; nuclear oncoprotein p53
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change
08-Sep-1997
ACCESSIONS S02192; S41149
REFERENCE S02192
#authors Soussi, T.; de Fromental, C.C.; Breugnot, C.; May, E.
#journal Nucleic Acids Res. (1988) 16:11384
#title Nucleotide sequence of a cDNA encoding the rat p53 nuclear
oncoprotein.
#cross-references M01D:89083585
#accession S02192
#molecule_type mRNA
#residues 1-391 ##label S0U
#cross-references EMBL:X13058; M1D:956828; PID:956829
REFERENCE S41149
#authors Hulla, J.E.; Schneider, R.P.
#journal Nucleic Acids Res. (1993) 21:713-717
#title Structure of the rat p53 tumor suppressor gene.
#accession S41149
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-173, 'W', 175-391 ##label HUL
#cross-references EMBL:L07909
#note the nucleotide sequence was submitted to the EMBL Data
Library, December 1992
GENETICS
#introns 25/2; 32/3; 123/3; 185/1; 259/2; 305/1; 329/3; 365/2
CLASSIFICATION
#superfamily cellular tumor antigen p53
#apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
FEATURE
174,177,236,240 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
390 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted

```

SUMMARY #length 391 #molecular-weight 43451 #checksum 7105  
Query Match 100.0%; Score 64; DB 2; Length 391;  
Best Local Similarity 100.0%; Pred. No. 1.57e-02;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 247 RPLIITL 255  
1 RPLIITL 9  
RESULT 13  
ENTRY #type complete  
TITLE tumor suppressor p53 - rabbit  
ORGANISM #formal\_name Oryctolagus cuniculus #common\_name domestic  
#alt\_name  
DATE 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change  
ACCESSIONS JC6193  
REFERENCE JC6193  
#authors Le Gass, F.; May, P.; Ronco, P.; de Fromental, C.C.  
#journal Gene (1997) 185:169-173  
#title cDNA cloning and immunological characterization of rabbit  
#accession JC6193  
#molecule\_type mRNA  
#residues 1-391 #label LBA  
#cross-references EMBL:X50592; NID:g1532043; PID:e194962; PID:g1532044  
GENETICS  
#gene p53  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS tumor  
SUMMARY #length 391 #molecular-weight 43435 #checksum 4367  
Query Match 100.0%; Score 64; DB 2; Length 391;  
Best Local Similarity 100.0%; Pred. No. 1.57e-02;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 246 RPLIITL 254  
1 RPLIITL 9  
RESULT 14  
ENTRY #type complete  
TITLE tumor suppressor protein p53 - Chinese hamster  
ORGANISM #formal\_name Cricetus griseus #common\_name Chinese hamster  
DATE 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change  
ACCESSIONS JC6176  
REFERENCE JC6176  
#authors Lee, H.; Larner, J.M.; Hamlin, J.L.  
#journal Gene (1997) 184:177-183  
#title Cloning and characterization of Chinese hamster p53 cDNA.  
#contents Liver  
#accession JC6176  
#molecule\_type mRNA  
#residues 1-393 #label LBE  
#cross-references GB:050395; NID:g1842229; PID:g1842230  
COMMENT This protein is a multimer, it plays the central role in a complex  
DNA damage-sensing network, it binds to replication factor and  
RNA-binding protein, and affects DNA replication, transcription,  
and recombination by protein/protein interactions.  
GENETICS  
#gene p53  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS liver; tumor  
SUMMARY #length 393 #molecular-weight 43362 #checksum 4043  
Query Match 100.0%; Score 64; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.57e-02;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 249 RPLIITL 257  
1 RPLIITL 9  
RESULT 15  
ENTRY #type complete  
TITLE cellular tumor antigen p53 - human  
ALTERNATE\_NAMES cellular phosphoprotein p53; oncoprotein p53; transformation  
suppressor p53; tumor suppressor p53  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 05-Oct-1988 #sequence\_revision 18-Nov-1994 #text\_change  
ACCESSIONS A55224; A43073; JT0436; S40773; S42669; A22837; A55060;  
A25397; B25397; S42452; S42453; I38082; I38083; I38084;  
I38085; I38086; I38087; I38088; I38089; I38090; I38091;  
I38092; I38093; A44905; I58354; I78850; S60153  
REFERENCE A25224  
#authors Lamb, P.; Crawford, L.  
#journal Mol. Cell. Biol. (1986) 6:1379-1385  
#title Characterization of the human p53 gene.  
#cross-references MUID:87064416  
#accession A25224  
#molecule\_type DNA  
#residues 1-393 #label LAM  
#cross-references EMBL:X01405; GB:M13121; GB:N00032; NID:g189460;  
PID:g386994  
REFERENCE JT0436  
#authors Buchman, V.L.; Chumakov, P.M.; Minkina, N.N.; Samarina, O.P.;  
Georgiev, G.P.  
#journal Gene (1988) 70:245-252  
#title A variation in the structure of the protein-coding region of  
the human p53 gene.  
#cross-references MUID:89106008  
#accession A43073  
#molecule\_type DNA  
#residues 1-393 #label BUC  
#note this 72-Arg allele appears to be about 5 times more  
frequent than the 72-Pro allele  
#accession JT0436  
#molecule\_type DNA  
#residues 1-71, 'P', 73-393 #label BU2  
#cross-references EMBL:M2898; NID:g189474  
#note this 72-Pro allele was found in both normal and  
malignant cell lines  
REFERENCE S40773  
#authors Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.  
#subission submitted to the EMBL Data Library, August 1990  
#accession S40773  
#molecule\_type DNA  
#residues 1-393 #label CHU  
#cross-references EMBL:X54156; NID:g35213; PID:g35214  
S42669  
REFERENCE S42669  
#authors Matlashevskii, G.; Lamb, P.; Plm, D.; Peacock, J.; Crawford,  
L.; Benchimol, S.  
#journal EMBO J. (1984) 3:3257-3262  
#title Isolation and characterization of a human p53 cDNA clone:  
expression of the human p53 gene.  
#accession S42669  
#molecule\_type mRNA  
#residues 101-393 #label MKI  
#cross-references EMBL:X01405; NID:g35215; PID:g642241  
REFERENCE A22837  
#authors Zakut-Houri, R.; Bienz-Tadnor, B.; Givol, D.; Oren, M.  
#journal EMBO J. (1985) 4:1251-1255  
#title Human p53 cellular tumor antigen: cDNA sequence and  
expression in COS cells.  
#cross-references MUID:85230577  
#accession A22837  
#molecule\_type mRNA  
#residues 1-71, 'P', 73-393 #label ZAK  
#cross-references EMBL:X02469; EMBL:M60950; NID:g35209; PID:g35210  
REFERENCE A55060

```

#authors Harlow, E.; Williamson, N.M.; Ralston, R.; Helfman, D.M.;
#journal Mol. Cell. Biol. (1985) 5:1601-1610
#title Molecular cloning and in vitro expression of a cDNA clone for
#accession A55060
#molecule_type mRNA
#residues 1-71, 'P', 73-272, 'H', 274-393 ##label HA3
#cross-references GB:K03199; NID:g189478; PID:g189479
#experimental_source clone p4-2, cell line A431
REFERENCE
#authors Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.;
#journal Mol. Cell. Biol. (1986) 6:4650-4656
#title Molecular basis for heterogeneity of the human p53 protein.
#cross-references MIM:87089326
#accession A25397
#molecule_type mRNA
#residues 1-78, 'T', 80-393 ##label HAR
#cross-references EMBL:M4694; NID:g339813; PID:g339814
#experimental_source clone p53-H-1, transformed hybridoma SV-80 cell
#accession B25397
#molecule_type mRNA
#residues 1-71, 'P', 73-78, 'T', 80-393 ##label HA2
#cross-references EMBL:M4695; NID:g339815; PID:g339816
#experimental_source clone p53-H-19, transformed hybridoma SV-80 cell
#accession S42452
#authors Matlshewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider,
#journal J. Crawford, L.V.
#title Primary structure polymorphism at amino acid residue 72 of
#accession S42452
#molecule_type mRNA: DNA
#residues 66-71, 'P', 73-79 ##label MK2
#experimental_source clone lambda C13
#note 72-Cys was also found, and appears to represent a
#accession S42453
#molecule_type mRNA: DNA
#residues 66-79 ##label MAT
#experimental_source clone J6K
REFERENCE
#authors Faircliff, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.;
#journal EMBO J. (1991) 10:2879-2887
#title p53 is frequently mutated in Burkitt's lymphoma cell lines.
#cross-references MIM:9200731
#accession I38082
#molecule_type mRNA
#residues 1-189, 'LISLSEKKEICVMSIMTETLEDDIVWCMGMSRLRLAL',
#experimental_source 'VPSTTTCTVTPAMAA' ##label F01
#cross-references EMBL:X60010; NID:g506432; PID:g506433
#note deletion of a C nucleotide causes a frameshift at
#accession I38083
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-192, 'R', 194-393 ##label F02
#cross-references EMBL:X60011; NID:g506434; PID:g506435
#accession I38084
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-393 ##label F03
#cross-references EMBL:X60012; NID:g506436; PID:g506437
#accession I38085
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-245, 'T', 247-393 ##label F04
#cross-references EMBL:X60013; NID:g506438; PID:g506439

```

```

#accession I38086
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-236, 'T', 238-393 ##label F05
#cross-references EMBL:X60014; NID:g506440; PID:g506441
#accession I38087
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-247, 'Q', 249-393 ##label F06
#cross-references EMBL:X60015; NID:g506442; PID:g506443
#accession I38088
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-71, 'P', 73-237, 'Y', 239-393 ##label F07
#cross-references EMBL:X60016; NID:g506444; PID:g506445
#accession I38089
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-247, 'Q', 249-393 ##label F08
#cross-references EMBL:X60017; NID:g506446; PID:g506447
#accession I38090
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-71, 'P', 73-162, 'H', 164-393 ##label F09
#cross-references EMBL:X60018; NID:g506448; PID:g506449
#accession I38091
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-212, 'Q', 214-393 ##label F10
#cross-references EMBL:X60019; NID:g506450; PID:g506451
#accession I38092
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-253, 'D', 255-393 ##label F11
#cross-references EMBL:X60020; NID:g506452; PID:g506453
#note all sequences submitted to the EMBL/GenBank/DBJ
#accession I38093
#status translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-393 ##label RE2
#cross-references EMBL:X54156; NID:g35213; PID:g35214
REFERENCE
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;
#journal Cancer Res. (1991) 51:5800-5805
#title p53 gene mutations in gastric cancer metastases and in
#experimental_source gastric cancer cell lines derived from metastases.
#accession A44905
#molecule_type DNA
#residues 246-247, 'W', 249-250 ##label YAM
#cross-references GB:S63157; NID:g237829; PID:g237830
#note sequence extracted from NCBI backbone (NCBIN:63157,
#accession I38093
#status translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-393 ##label RE2
#cross-references EMBL:X54156; NID:g35213; PID:g35214
REFERENCE
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;
#journal Cancer Res. (1991) 51:5800-5805
#title p53 gene mutations in gastric cancer metastases and in
#experimental_source gastric cancer cell lines derived from metastases.
#accession A44905
#molecule_type DNA
#residues 246-247, 'W', 249-250 ##label YAM
#cross-references GB:S63157; NID:g237829; PID:g237830
#note sequence extracted from NCBI backbone (NCBIN:63157,

```

Note: remainder of annotations omitted.

```

Query Match 100.0%; Score 64; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 1,57e-02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 249 RPLTITTL 257
1 RPLTITTL 9

```

Search completed: Fri Sep 11 13:40:55 1998

Sun Sep 13 10:55:45 1998

Job time : 21 secs.

US-08-452-843-17.rpr

$(\mathbf{w}_l)$ 

```
Run on:      Fri Sep 11 13:41:13 1998;      MasPar time 2.42 Seconds
Tabular output not generated.              93.368 Million cell updates/sec
```

|                |                           |
|----------------|---------------------------|
| Title:         | >US-08-452-843-17         |
| Description:   | (1-9) from US08452843.pep |
| Perfect Score: | 64                        |
| Sequence:      | 1 RPIILITTL 9             |

Scoring table: PAM 150

Searched: 69111 seqs, 25083644 residues

```
Post-processing: Minimum Match 0%
Listing first 45 summaries
```

```
Database: swiss-prot35
          1:swiss1'
```

Statistics: Mean 24.140; Variance 25.966; scale 0.930

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| No. | Score | Query | Match | Length | DB         | ID | Description            | Pred.    | No. |
|-----|-------|-------|-------|--------|------------|----|------------------------|----------|-----|
| 1   | 64    | 100.0 | 207   | 1      | P53_EOVAS  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 2   | 64    | 100.0 | 276   | 1      | P53_CAFPA  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 3   | 64    | 100.0 | 280   | 1      | P53_HOMSE  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 4   | 64    | 100.0 | 314   | 1      | P53_SPEBE  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 5   | 64    | 100.0 | 383   | 1      | P53_XEMIA  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 6   | 64    | 100.0 | 373   | 1      | P53_BRBRE  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 7   | 64    | 100.0 | 382   | 1      | P53_SHEEP  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 8   | 64    | 100.0 | 386   | 1      | P53_BOVIN  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 9   | 64    | 100.0 | 380   | 1      | P53_MOUSE  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 10  | 64    | 100.0 | 391   | 1      | P53_RABIT  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 11  | 64    | 100.0 | 391   | 1      | P53_RAT    |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 12  | 64    | 100.0 | 393   | 1      | P53_HUMAN  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 13  | 64    | 100.0 | 393   | 1      | P53_CRGR   |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 14  | 64    | 100.0 | 393   | 1      | P53_CERAE  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 15  | 64    | 100.0 | 396   | 1      | P53_MESAU  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 16  | 64    | 100.0 | 396   | 1      | P53_SALIR  |    | CELLULAR TUMOR ANTIGEN | 8.666-04 |     |
| 17  | 59    | 92.2  | 351   | 1      | P53_ORLIA  |    | CELLULAR TUMOR ANTIGEN | 1.655-02 |     |
| 18  | 59    | 92.2  | 357   | 1      | P53_CHICK  |    | CELLULAR TUMOR ANTIGEN | 1.655-02 |     |
| 19  | 58    | 90.6  | 386   | 1      | P53_FELCA  |    | CELLULAR TUMOR ANTIGEN | 2.938-00 |     |
| 20  | 51    | 79.7  | 366   | 1      | P53_PLAIF  |    | CELLULAR TUMOR ANTIGEN | 1.365-00 |     |
| 21  | 50    | 78.1  | 328   | 1      | YHHD_ERMCH |    | HYPOTHETICAL 36.0 KD P | 2.286-00 |     |
| 22  | 50    | 78.1  | 522   | 1      | CP4_RAT    |    | CYTICHROME P450 IV4 P  | 2.286-00 |     |
| 23  | 48    | 75.0  | 397   | 1      | YXDA_BACGS |    | HYPOTHETICAL 43.7 KD P | 6.316-00 |     |

|    |    |      |      |   |             |                          |          |
|----|----|------|------|---|-------------|--------------------------|----------|
| 45 | 24 | 48   | 532  | 1 | FM05_HUMAN  | DIETHYLANILINE MONOOX    | 6.31e+00 |
| 25 | 48 | 75.0 | 532  | 1 | FM05_CAVPO  | DIETHYLANILINE MONOOX    | 6.31e+00 |
| 26 | 47 | 73.4 | 464  | 1 | KTR4_YEAST  | PROBABLE MANNOSYLTRANS   | 1.04e+07 |
| 27 | 46 | 71.9 | 100  | 1 | VG10_HSVB   | HYPOTHETICAL GENE 10 P   | 1.68e+07 |
| 28 | 46 | 71.9 | 439  | 1 | BRNO_ECOLI  | BRANCHED CHAIN AMINO A   | 1.68e+07 |
| 29 | 46 | 71.9 | 439  | 1 | BRNO_SALTY  | BRANCHED CHAIN AMINO A   | 1.68e+07 |
| 30 | 46 | 71.9 | 698  | 1 | SGS1_YEAST  | SPIRULATION PROTEIN GS   | 1.68e+07 |
| 31 | 45 | 70.3 | 222  | 1 | OLF4_MOUSE  | OLFACTORY RECEPTOR-LIK   | 2.72e+00 |
| 32 | 45 | 70.3 | 226  | 1 | PIRE_YEAST  | OROTATE PHOSPHORIBOSYL   | 2.72e+00 |
| 33 | 45 | 70.3 | 227  | 1 | PIRX_YEAST  | OROTATE PHOSPHORIBOSYL   | 2.72e+00 |
| 34 | 45 | 70.3 | 323  | 1 | YKJ2_GABEL  | HYPOTHETICAL 36.9 KD P   | 2.72e+00 |
| 35 | 45 | 70.3 | 323  | 1 | QOX2_BASCU  | QUINOL OXIDASE POLYPEP   | 2.72e+00 |
| 36 | 45 | 70.3 | 452  | 1 | PTCC_ECOLI  | PTS SYSTEM, CELLOBIOSE   | 2.72e+00 |
| 37 | 45 | 70.3 | 604  | 1 | YFIC_BASCU  | HYDROPHETICAL ABC TRANS  | 2.72e+00 |
| 38 | 45 | 70.3 | 868  | 1 | WGBL_VZVD   | GLYCOPROTEIN B PRECURS   | 2.72e+00 |
| 39 | 44 | 68.8 | 119  | 1 | FRDD_ECOLI  | IMMUNITE REDUCTASE I3    | 4.34e+01 |
| 40 | 44 | 68.8 | 141  | 1 | IMMM_ECOLI  | IMMUNITY PROTEIN FOR C   | 4.34e+01 |
| 41 | 44 | 68.8 | 154  | 1 | FLIIL_ECOLI | FLAGELLAR FLII PROTEIN   | 4.34e+01 |
| 42 | 44 | 68.8 | 348  | 1 | NAM2_SCHPO  | PEROMYONE P-FACTOR REC   | 4.34e+01 |
| 43 | 44 | 68.8 | 414  | 1 | FTSW_ECOLI  | CITRATE DIVISION PROTEIN | 4.34e+01 |
| 44 | 44 | 68.8 | 431  | 1 | CITL_ECOLI  | CITRATE-PROTON SYMPORT   | 4.34e+01 |
| 45 | 44 | 68.8 | 1458 | 1 | PA2R_RABIT  | 160 KD SECRETORY PROSP   | 4.34e+01 |

## ALIGNMENTS

| RESULT                | 1   | STANDARD:    | PRI:                | 207 AA.                                  |
|-----------------------|---|--------------|---------------------|--|
| ID                    | P53_EOUAS   |              |                     |  |
| AC                    | 029480:   |              |                     |  |
| DT                    | 01-NOV-1997 (REL. 35, CREATED)  |              |                     |  |
| DT                    | 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)   |              |                     |  |
| DT                    | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)   |              |                     |  |
| DE                    | CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  |              |                     |  |
| GN                    | TP53.   |              |                     |  |
| OS                    | EOUAS ASINUS (DONKEY).  |              |                     |  |
| OC                    | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  |              |                     |  |
| OC                    | ETHIRIA; PERISSODACTYLA.  |              |                     |  |
| RN                    | [1]   |              |                     |  |
| RP                    | SEQUENCE FROM N.A.  |              |                     |  |
| RX                    | MEDLINE: 96342529.  |              |                     |  |
| RA                    | NASIR I., REID S.W.;  |              |                     |  |
| RL                    | DNA SEQ. 6,61-63(1995).   |              |                     |  |
| CC                    | -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND P53 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION (BY SIMILARITY). |              |                     |  |
| CC                    | -1- SUBCELLULAR LOCATION: NUCLEAR.  |              |                     |  |
| CC                    | -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.  |              |                     |  |
| CC                    | -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  |              |                     |  |
| DR                    | EMBL: U26741; G1020153; .   |              |                     |  |
| DR                    | PROSITE: P500348; P53; 15.  |              |                     |  |
| KW                    | ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  |              |                     |  |
| KW                    | NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  |              |                     |  |
| FT                    | DOMAIN  | 1            |                     |  |
| FT                    | NON_TER   | 1            |                     |  |
| FT                    | NON_TER   | 187          | 199                 | NUCLEAR LOCALIZATION SIGNAL (POTENTIAL). |
| FT                    | NON_TER   | 207          | 207                 |  |
| FT                    | SEQUENCE  | 207 AA:      | 23428 MW:           | OFBAP9C1 CRC32;                          |
| Query Match           |   | 100.0%;      | Score 64;           | DB 1; Length 207;                        |
| Best Local Similarity |   | 100.0%;      | Pred. No. 8,66e-04; |  |
| Matches               | 9;  | Conservative | 0;                  | Mismatches 0; Indels 0; Gaps 0           |
| Db                    | 125 RPL1ITL 133   |              |                     |  |
|                       |   |              |                     |  |
|                       | 1 RPL1ITL 9   |              |                     |  |

RESULT 2  
ID P53\_CANFA STANDARD: PRT: 276 AA.  
AC Q29537;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53.  
OS CANIS FAMILIARIS (DOG).  
OC EUKARYOTA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
OC EUTHERIA: CARNIVORA.  
RN [1].  
RP SEQUENCE FROM N.A.  
RC STRAIN-BEAGLE;  
RX MEDLINE: 95323915.  
RA KRAEGL S.A., PAZZI K.A., MADEWELL B.R.;  
RL CANCER LETT. 92:181-186(1995).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC EMBL: S77819; G1000577;  
CC PROSITE: PS00348; P53; 1.  
CC ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
CC KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
CC FT NON\_TER 1 1  
CC FT DOMAIN 1 35 ASP/GLU-RICH (ACIDIC).  
CC FT NON\_TER 276 276  
CC FT SEQUENCE 276 AA: 30466 MW: 8697AE44 CRC32:  
SQ

Query Match 100.0%; Score 64; DB 1; Length 276;  
Best Local Similarity 100.0%; Pred. No. 8.66e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 213 RLITITL 221  
QY 1 RLITITL 9

RESULT 3  
ID P53\_HORSE STANDARD: PRT: 280 AA.  
AC P79892; Q29481;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53 OR P53.  
OS EQUUS CABALLUS (HORSE).  
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
OC EUTHERIA: PERISSODACTYLA.  
RN [1].  
RP SEQUENCE OF 1-263 FROM N.A.  
RC TISSUE=SPLEEN;  
RX MEDLINE: 97070350.  
RA PAZZI K.A., KRAEGL S.A., GRIFFEY S.M., THEON A.P., MADEWELL B.R.;  
RL CANCER LETT. 107:123-130(1996).  
RN [2].  
RN SEQUENCE OF 76-280 FROM N.A.  
RX MEDLINE: 96293865.  
RA NASIR L., REID S.W.;

RL DNA SEQ. 6:185-187(1996).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC EMBL: S83123; G1836145;  
CC DR EMBL: U37120; G1389675;  
CC PROSITE: PS00348; P53; 1.  
CC ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
CC KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
CC FT NON\_TER 1 1  
CC FT DOMAIN 1 1  
CC FT CONFLICT 79 274 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC FT CONFLICT 79 79 T -> A (IN REF. 2).  
CC FT CONFLICT 83 83 L -> M (IN REF. 2).  
CC FT CONFLICT 111 111 A -> V (IN REF. 2).  
CC FT CONFLICT 138 138 G -> A (IN REF. 2).  
CC FT NON\_TER 280 280  
CC FT SEQUENCE 280 AA: 30985 MW: B494F872 CRC32:  
SQ

Query Match 100.0%; Score 64; DB 1; Length 280;  
Best Local Similarity 100.0%; Pred. No. 8.66e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 200 RLITITL 208  
QY 1 RLITITL 9

RESULT 4  
ID P53\_SPEBE STANDARD: PRT: 314 AA.  
AC Q64662;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53.  
OS SPERMOPHILUS BEECHEYI (BEECHY GROUND SQUIRREL).  
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
OC EUTHERIA: RODENTIA.  
RN [1].  
RP SEQUENCE FROM N.A.  
RC TISSUE=THYMUS;  
RX MEDLINE: 95007566.  
RA RYKINA M.B., CULLEN J.M., ROBINSON W.S., MARION P.L.;  
RL CANCER RES. 54:5430-5437(1994).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC EMBL: U43902; G1165312;  
CC PROSITE: PS00348; P53; 1.  
CC DR



KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION.

FT NON\_TER 1 1 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 289 301  
 FT NON\_TER 314 314  
 SQ SEQUENCE 314 AA: 34618 MW: D07F433B CRC32;

Query Match 100.0%; Score 64; DB 1; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 8.66e-04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 227 RPILITITL 235  
 |||||  
 RPILITITL 9

RESULT 5 STANDARD; PRT: 363 AA.

ID P53\_XENLA  
 AC P07193;

DT 01-APR-1988 (REL. 07, CREATED)  
 DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53.

GN TP53.  
 OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.

RN [1]  
 RP SEQUENCE FROM N.A.

RA MEDLINE: 88143684.  
 RA SOUSSEI T., DE FROMENTEL C.C., MECHALI M., MAY P., KRESS M.;  
 RL ONCOGENE 1:71-78(1987).

RN [2]  
 RP SEQUENCE FROM N.A.

RA MEDLINE: 94134403.  
 RA HOEVER M., CLEMENT J.H., WEDLICH D., MONTENARH M., KNOCHEL W.;  
 RL ONCOGENE 9:109-120(1994).

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL

CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A

TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF

THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF

BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 EXPRESSION (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: UBICUITOUS.

CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

DR EMBL: M36962; G214640; -

DR EMBL: X05191; G64962; -

DR EMBL: X77546; G468514; -

DR EMBL: S68353; G545102; -

DR PIR: A29376; A29376.

DR HSSP: P04637; 1PES.

DR PROSITE: PS00348; P53; 1.

KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.

FT DOMAIN 281 293 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 362 362 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 52 52 T->S (IN REF. 2).  
 FT CONFLICT 71 71 MISSING (IN REF. 2).  
 FT CONFLICT 296 296 MISSING (IN REF. 2).

SQ SEQUENCE 363 AA: 40692 MW: 7507D796 CRC32;

Query Match 100.0%; Score 64; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 8.66e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 224 RPILITITL 232  
 |||||  
 RPILITITL 9

RESULT 6 STANDARD; PRT: 373 AA.

ID P53\_BRARE  
 AC P79734;

DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53.

GN TP53.

OS BRACHYDANTIO RERIO (ZEBRAFISH) (ZEBRA DANIO).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA.

RN [1]  
 RP SEQUENCE FROM N.A.

RA BAILEY G.S., CHENG R., FORD B.L., O'NEAL P.E., MATHEWS C.Z.,  
 RA BRADFORD C.S., THONGTAN T., BARNES D.W., HENDRICKS J.D.;  
 RL MOL. MAR. BIOL. BIOTECHNOL. 6:88-97(1997).

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL

CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A

TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF

THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF

BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 EXPRESSION (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

DR EMBL: U60804; G1778019; -

DR PROSITE: PS00348; P53; 1.

KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.

FT DOMAIN 280 296 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 372 372 PHOSPHORYLATION (BY SIMILARITY).

SQ SEQUENCE 373 AA: 41899 MW: 706A4B9C CRC32;

Query Match 100.0%; Score 64; DB 1; Length 373;  
 Best Local Similarity 100.0%; Pred. No. 8.66e-04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 217 RPILITITL 225  
 |||||  
 RPILITITL 9

RESULT 7 STANDARD; PRT: 382 AA.

ID P53\_SHEEP  
 AC P51664;

DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53.

GN TP53.

OS OVIS ARIES (SHEEP).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA TISSUE-BLOOD.  
 RA MEDLINE: 95352828.  
 RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
 RL DNA SEQ. 5:255-259(1992).

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL

CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A

TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 EXPRESSION.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: X61705: G602357; -  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 66 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 300 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD RES 381 381 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 382 AA: 42809 MW: 0CB99A00 CRC32:  
 Query Match 9 100.0% Score 64; DB 1; Length 382;  
 Best Local Similarity 100.0% Pred. No. 8,66e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 238 RPLITITL 246  
 QY 1 RPLITITL 9  
 RESULT 8  
 ID P53\_BOVIN STANDARD; PRT; 386 AA.  
 AC 029628;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS BOS TAURUS (BOVINE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RA MEDLINE: 95352829.  
 RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
 RL DNA SEQ. 5:261-264(1995).  
 RN [2]  
 RP SEQUENCE OF 13-386 FROM N.A.  
 RC STRAIN-HOLSTEIN; TISSUE-THYMUS;  
 RX MEDLINE: 96401400.  
 RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;  
 RL VET. IMMUNOL. IMMUNOPATHOL. 52:53-63(1996).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 BAX AND P53 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: X61704: G602333; -  
 DR EMBL: X49825: G1729419; -  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 380 380 R -> T (IN REF. 2).  
 SQ SEQUENCE 386 AA: 43255 MW: 0322BF3D CRC32:  
 Query Match 9 100.0% Score 64; DB 1; Length 386;  
 Best Local Similarity 100.0% Pred. No. 8,66e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 242 RPLITITL 250  
 QY 1 RPLITITL 9  
 RESULT 9  
 ID P53\_MOUSE STANDARD; PRT; 390 AA.  
 AC P02340;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR TRP53 OR P53.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE: 85027173.  
 RA BIENZ B., ZAKUT-HOURI R., GIVOL D., OREN M.;  
 RL EMBL J. 3:2179-2183(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84068204.  
 RA ZAKUT-HOURI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIVOL D.;  
 RL NATURE 306:594-597(1983).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84272240.  
 RA JENKINS J.R., RUDGE K., REDMOND S., MADE-EVANS A.;  
 RL NUCLEIC ACIDS RES. 12:5609-5626(1984).  
 RN [4]  
 RP SEQUENCE FROM N.A. (CLONES PCDS3; P53-M11 AND P53-M8).  
 RX MEDLINE: 87064640.  
 RA ARAI N., NOMURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.,  
 RA ROTTER V.;  
 RL MOL. CELL. BIOL. 6:3232-3239(1986).  
 RN [5]  
 RP SEQUENCE OF 222-258 FROM N.A.  
 RX MEDLINE: 92115342.  
 RA BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BRENNER R.,  
 RA BALMAIN A.;  
 RL ONCOGENE 6:2363-2369(1991).  
 RN [6]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE: 86149247.  
 RA SAMD A., ANDERSON C.W., CARROLL R.B.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:897-904(1986).  
 RN [7]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE: 91006019.  
 RA WEEK D.W., SIMON S., KIKKAWA U., ECKHART W.;  
 RL EMBL J. 9:3253-3260(1990).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 BAX AND P53 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: X00876: G871421; JOINED.  
 DR EMBL: X00877: G871421; JOINED.  
 DR EMBL: X00878: G871421; JOINED.  
 DR EMBL: X00879: G871421; JOINED.  
 DR EMBL: X00880: G871421; JOINED.

DR EMBL: X00881; 6871421; JOINED.  
 DR EMBL: X00882; 6871421; JOINED.  
 DR EMBL: X00883; 6871421; JOINED.  
 DR EMBL: X00884; 6871421; JOINED.  
 DR EMBL: X00885; 6871421; JOINED.  
 DR EMBL: K01700; G200203; -  
 DR EMBL: K01237; G53576; -  
 DR EMBL: X00741; G53571; -  
 DR EMBL: M13872; G200199; -  
 DR EMBL: M13873; G200201; -  
 DR EMBL: M13874; G200203; -  
 DR EMBL: S77930; G243255; -  
 DR EMBL: S77930; G243255; -  
 DR PIR: A02684; DMM53.  
 DR PIR: A22739; A22739.  
 DR PIR: S38822; S38822.  
 DR HSSP: P04637; IPES.  
 DR TRANSFAC: T01806; -  
 DR MGD: MGI:98834; TRP53.  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS; DISEASE MUTATION.  
 FT DOMAIN 1 75  
 FT DOMAIN 76 150  
 FT DOMAIN 276 390  
 FT DOMAIN 308 320  
 FT MOD.RES 312 312  
 FT MOD.RES 389 389  
 FT VARIANT 135 135  
 FT VARIANT 168 168  
 FT CONFLICT 48 48  
 FT CONFLICT 79 81  
 SQ SEQUENCE 390 AA; 43458 MM; 8943DB93 CRC32;  
 PVA -> OM (IN REF. 3).  
 Query Match  
 Best Local Similarity 100.0%; Score 64; DB 1; Length 390;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 246 RPLTITL 254  
 QY 1 RPLTITL 9

RESULT 10  
 ID P53-RABIT STANDARD: PRT: 391 AA.  
 AC 095330;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS ORYCTOLAGUS CUNICULUS (RABBIT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; LAGOMORPHA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NEW ZEALAND;  
 RX MEDLINE: 97208869.  
 RA LE GOAS F., MAY P., RONCO P., CARON DE FROMENTEL C.;  
 RL GENE 185:169-173(1997).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND BAX ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION (BY SIMILARITY).  
 CC SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY

CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: X00592; E194962; -  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 70  
 FT DOMAIN 308 321  
 FT MOD.RES 390 390  
 SQ SEQUENCE 391 AA; 43435 MM; 30A36172 CRC32;  
 PVA -> OM (IN REF. 3).  
 Query Match  
 Best Local Similarity 100.0%; Score 64; DB 1; Length 391;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 246 RPLTITL 254  
 QY 1 RPLTITL 9

RESULT 11  
 ID P53-RAT STANDARD: PRT: 391 AA.  
 AC P10361; 009168;  
 DT 01-MAR-1989 (REL. 10, CREATED)  
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR P53.  
 OS RATUUS NOVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE: 89083585.  
 RA SOUSSE T.;  
 RL NUCLEIC ACIDS RES. 16:11384-11384(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE: 93181268.  
 RA HULLA J.E., SCHNEIDER R.P.;  
 RL NUCLEIC ACIDS RES. 21:713-717(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RA MATHUPALA S.P.;  
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND BAX ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: X13058; G56829; -  
 DR EMBL: L07910; G205952; -  
 DR EMBL: L07904; G205952; JOINED.  
 DR EMBL: L07905; G205952; JOINED.  
 DR EMBL: L07906; G205952; JOINED.  
 DR EMBL: L07907; G205952; JOINED.  
 DR EMBL: L07908; G205952; JOINED.  
 DR EMBL: L07909; G205952; JOINED.  
 DR EMBL: U90328; G1938365; -  
 DR PIR: S02192; S02192.  
 DR HSSP: P04637; IPES.

DR PROSITE: PS00348; P53; 1.  
KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
FT NUCLEAR PROTEIN: PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 1 76 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 77 151 HYDROPHOBIC.  
FT DOMAIN 277 391 HIGHLY BASIC AND MAY BE INVOLVED IN  
INTERACTION WITH DNA.  
FT DOMAIN 309 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD RES 390 390 PHOSPHORYLATION (BY SIMILARITY).  
FT VARIANT 103 103  
FT VARIANT 256 256  
FT \*CONFLICT 174 174 E -> S.  
FT \*CONFLICT 174 174 C -> W (IN REF. 2).  
SQ \*SEQUENCE 391 AA; 43451 MM; ED114C18 CRC32;  
Query Match 100.0%; Score 64; DB 1; Length 391;  
Best Local Similarity 100.0%; Pred. No. 8,666-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 247 RPIITITL 255  
Qy 1 RPIITITL 9  
RESULT 12  
ID P53\_HUMAN STANDARD; PRT; 393 AA.  
AC P04637;  
DT 13-AUG-1987 (REL. 05, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
GN TP53.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85230577.  
RA ZAKUT-HOURI R., BIENZ-TADMOR B., GIOVL D., OREN M.;  
EMBO J. 4:1251-1255(1985).  
RL [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 87064416.  
RA LAMB P., CRAWFORD L.;  
MOL. CELL. BIOL. 6:1379-1385(1986).  
RL [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85267676.  
RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;  
MOL. CELL. BIOL. 5:1601-1610(1985).  
RL [4]  
RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
RX MEDLINE; 87089826.  
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
ROTTER V.;  
MOL. CELL. BIOL. 6:4650-4656(1986).  
RL [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89108008.  
RA BUCHMAN V.L., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P.,  
GENE 70:245-252(1988).  
RL [6]  
RP SEQUENCE OF 101-393 FROM N.A.  
RX MEDLINE; 85126934.  
RA MATLASHENSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
BENCHMOL S.;  
EMBO J. 3:3257-3262(1984).  
RL [7]  
RP NUCLEAR LOCALIZATION SIGNAL.  
RX MEDLINE; 90191730.  
RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
ONCOGENE 5:423-426(1990).  
RL [8]  
RP STRUCTURE BY NMR OF 319-360.  
RX MEDLINE; 94294808.  
RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
APPELLA E., GROENENORN A.M.;  
SCIENCE 265:386-391(1994).  
RL [9]  
RP STRUCTURE BY NMR OF 325-355.  
RX MEDLINE; 95292092.  
RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
NAT. STRUCT. BIOL. 1:877-890(1994).  
RL [10]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
RX MEDLINE; 94294806.  
RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
SCIENCE 265:346-355(1994).  
RL [11]  
RP REVIEW.  
RX MEDLINE; 94090335.  
RA HARRIS C.C.;  
SCIENCE 262:1980-1981(1993).  
RL [12]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 91289156.  
RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
SCIENCE 253:49-53(1991).  
RL [13]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 96271983.  
RA DE VRIES E.M.G., RICE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
LIAO D., SOSSI T., KOVACH J.S., SOMMER S.S.;  
HUM. MUTAT. 7:202-213(1996).  
RL [14]  
RP VARIANT ARG-72.  
RX MEDLINE; 91153807.  
RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
HUM. GENET. 86:369-370(1991).  
RL [15]  
RP VARIANT LFS THR-133.  
RX MEDLINE; 92034774.  
RA LAM J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
CANCER RES. 51:6385-6387(1991).  
RL [16]  
RP VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
RX MEDLINE; 91057657.  
RA WALKIN D., LI F.P., STRONG L.C., FROMENT J.F. JR., NELSON C.E.,  
KIM D.H., KASSEL J., GRAY M.A., BISCHOFF F.Z., TALINSKY M.A.,  
FRIEND S.H.;  
SCIENCE 250:1233-1238(1990).  
RL [17]  
RP VARIANT LFS ASP-245.  
RX MEDLINE; 91080929.  
RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
NATURE 348:747-749(1990).  
RL [18]  
RP VARIANT LFS LEU-272.  
RX MEDLINE; 92147883.  
RA FELIX C.A., NAD M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
KNOTSEN T., MINNA J.D.;  
J. CLIN. INVEST. 89:640-647(1992).  
RL [19]  
RP VARIANTS LFS HTS-273 AND VAL-325.  
RX MEDLINE; 92228023.  
RA WALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
GEBHARDT M.C., ANDERSEN T.I., BORRESSEN A.-L., LI F.P., GABBER J.,  
STRONG L.C.;  
NEW ENGL. J. MED. 326:1309-1315(1992).  
RL [20]  
RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
RX MEDLINE; 90295284.  
RA BARTER J., IGGO R., GANNON J., LANE D.P.;  
ONCOGENE 5:893-899(1990).  
RL [21]

RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
 RX MEDLINE: 91017544.  
 RA RODRIGUES N.R., ROWAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
 RA GANNON J.V., LANE D.P.,  
 RL PROC. NATL. ACAD. SCI. U.S.A. 87:7555-7558(1990).  
 RN [22]  
 RP VARIANTS COLORECTAL CANCER MUTATIONS.  
 RX MEDLINE: 91282784.  
 RA ISHIOKA C., SATO T., GAMOH M., SUZUKI T., SHIBATA H., KANAMARU R.,  
 RA WAKUI A., YAMAZAKI T.,  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 177:901-906(1991).  
 RN [23]  
 RP VARIANTS OESOPHAGUS TUMORS L-152; A-155; H-175; F-176 AND H-273.  
 RX MEDLINE: 91330175.  
 RA CASSON A.G., MUKHOPADHYAY T., CLEARY K.R., RO J.Y., LEVIN B.,  
 RA ROTH J.A.,  
 RL CANCER RES. 51:4495-4499(1991).  
 RN [24]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN CHINA.  
 RX MEDLINE: 91187113.  
 RA HSU I.C., METCALF R.A., SUN T., WELSH J.A., WANG N.J., HARRIS C.C.,  
 RL NATURE 350:427-428(1991).  
 RN [25]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN SOUTH AFRICA.  
 RX MEDLINE: 91187114.  
 RA BRESSAC B., KEM M., WANDS J., OZTURK M.,  
 RL NATURE 350:429-431(1991).  
 RN [26]  
 RP VARIANTS IN ANOGENITAL CARCINOMAS.  
 RX MEDLINE: 93010989.  
 RA CROOK T., VOUSDEN K.H.,  
 RL EMO J. 11:3935-3940(1992).  
 RN [27]  
 RP VARIANT WITH DUPLICATION OF PRO-177-HIS-178.  
 RX MEDLINE: 93265016.  
 RA BHATIA K., GUTIERREZ M.I., MAGRATH I.T.,  
 RL HUM. MOL. GENET. 1:207-208(1992).  
 RN [28]  
 RP VARIANTS IN BURKITT'S LYMPHOMAS.  
 RX MEDLINE: 93064692.  
 RA DUTRU A., DEBUIRE B., ROMANO J.W., EHRHART J.C., FISCELLA M., MAY E.,  
 RA APPELBA E., MAY P.,  
 RL ONCOGENE 7:2161-2167(1992).  
 RN [29]  
 RP VARIANT NASOPHARYNGEAL CARCINOMA THR-280.  
 RX MEDLINE: 92335329.  
 RA SUN Y., HEGAMER G., HENG Y.-J., HILDESHEIM A., CHEN J.-Y., CAO Y.,  
 RA YAO K.-T., COLEBURN N.H.,  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:6516-6520(1992).  
 RN [30]  
 RP VARIANTS IN COLON TUMORS.  
 RX MEDLINE: 93330562.  
 RA HAMELIN R., BEGO N., LAURENT-PUIG P., VIDAUD M., THOMAS G.,  
 RL ONCOGENE 8:2213-2220(1993).  
 RN [31]  
 RP CHARACTERIZATION OF VARIANT ALA-143.  
 RX MEDLINE: 94283378.  
 RA ZHANG W., GUO X.-Y., HU G.-Y., LIU W.-B., SHAY J.W., DEISSEROTH A.B.,  
 RL EMO J. 13:2535-2544(1994).  
 RN [32]  
 RP VARIANTS LFS HIS-175; ARG-193; GLN-248; CYS-273 AND TYR-275.  
 RX MEDLINE: 95193787.  
 RA FREBOURG T., BARBIER N., YAN Y.-X., GARBER J.E., DREYUS M.,  
 RA FRAUMONT J.F., JR., LI F.P., FRIEND S.H.,  
 RL AM. J. HUM. GENET. 56:608-615(1995).  
 RN [33]  
 RP VARIANT LFS HIS-175.  
 RX MEDLINE: 96423319.  
 RA VARLEY J.M., MCGROWN G., THORNCROFT M., TRICKER K.J., TEARE M.D.,  
 RA SANTIBANEZ-KOEF M.F., HOULSTON R.S., MARTIN J., BIRCH J.M.,  
 RL J. MED. GENET. 32:942-945(1995).  
 RN [34]

RP VARIANTS BA PHE-176; SER-245; TRP-248; TRP-282 AND GLN-286.  
 RX MEDLINE: 96233927.  
 RA AUDREZZE M.-P., ROBASKIEWICZ M., MERCIER B., NOUSAUD J.-B.,  
 RA HARDY E., BAIL J.-P., VOLANT A., LOZAC'H P., GOEROU H., FEREC C.,  
 RL HUM. MUTAT. 7:109-113(1996).  
 RN [35]  
 Note: remainder of annotations omitted.

Query Match 100.0%; Score 64; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8.66e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 249 RPLTITL 257  
 |||||||  
 QY 1 RPLTITL 9

RESULT 13  
 ID P53.CRIGR STANDARD; PRT; 393 AA.  
 AC 009185; 064397; P97258; P97788;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR P53.  
 OS CRICETULUS GRISEUS (CHINESE HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CHANG W., MI L.J., BOORSTEIN R.J.,  
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 97183659.  
 RA LEE H., LARNER J.M., HAMLIN J.L.,  
 RL GENE 184:177-183(1997).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG. TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC DR EMBL: Y08900; E303876; -;  
 CC DR EMBL: Y08901; E303863; -;  
 CC DR EMBL: U50395; G1842230; -;  
 CC DR PROSITE: PS00348; P53; 1.  
 CC KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 CC KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 CC FT DOMAIN 1 74 ASP/GLU-RICH (ACIDIC).  
 CC FT DOMAIN 75 150 HYDROPHOBIC.  
 CC FT DOMAIN 316 390 HIGHLY BASIC AND MAY BE INVOLVED IN  
 CC FT DOMAIN 311 323 INTERACTION WITH DNA.  
 CC FT MOD.RES 392 392 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT VARIANT 133 133 PHOSPHORYLATION (BY SIMILARITY).  
 CC FT VARIANT 135 135 L -> Q (IN CELL LINE V79-4).  
 CC FT CONFLICT 103 103 C -> W (IN CELL LINE V79-4).  
 CC FT SEQUENCE 393 AA; 43378 MW; 402EB149 CRC32;

Query Match 100.0%; Score 64; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8.66e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 249 RPLITITL 257  
 |||||  
 QY 1 RPLITITL 9

RESULT 14  
 ID P53 CERAE STANDARD: PRT: 393 AA.

AC P13481.  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.

OS CERCOPTHECUS AETHIOPS (GREEN MONKEY) (GRIVET)  
 OC EURAROTIA, MEZAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE-LIVER;  
 RX MEDLINE: 90045967.

RA RIGAUDY P., ECKHARDT W.;  
 RL NUCLEIC ACIDS RES. 17:8375-8375(1989).

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.

CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

CC EMBL: X16384; G23796;  
 CC PIR: S06594; S06594.

DR HSSP: P04637; 10LG.

DR PROSITE: PS00348; P53; 1.  
 DR ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;

KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.

FT DOMAIN 1 68 ASP/GLU-RICH (ACIDIC).

FT DOMAIN 81 150 HYDROPHOBIC.

FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN

FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

FT MOD RES 392 392 PHOSPHORYLATION (BY SIMILARITY).

SO SEQUENCE 393 AA; 43696 MW; BBE7DC62 CRC32;

DB 249 RPLITITL 257  
 |||||  
 QY 1 RPLITITL 9

RESULT 15  
 ID P53 MESAU STANDARD: PRT: 396 AA.

AC 000366; P97276;  
 DT 01-DEC-1992 (REL. 24, CREATED)  
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.

OS MESOCRICETUS AURATUS (GOLDEN HAMSTER).

OC EURAROTIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-SYRIAN; TISSUE-KIDNEY;  
 RX MEDLINE: 92210007.  
 RA LIEGROS Y., MCINTYRE P., SOUSSI T.;  
 RL GENE 112:247-250(1992).

RN [2]  
 RP SEQUENCE FROM N.A.

RA HOU E.W., WISEMAN R.;  
 RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.

CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

CC EMBL: M75144; G191415;  
 CC PIR: JH0633; JH0633.

DR HSSP: P04637; 1PES.

DR PROSITE: PS00348; P53; 1.  
 DR ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;

KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.

FT DOMAIN 1 77 ASP/GLU-RICH (ACIDIC).

FT DOMAIN 319 393 HYDROPHOBIC.

FT DOMAIN 314 326 HIGHLY BASIC AND MAY BE INVOLVED IN

FT MOD RES 395 395 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

FT CONFLICT 188 188 PHOSPHORYLATION (BY SIMILARITY).

SO SEQUENCE 396 AA; 43631 MW; C2668ADE CRC32;

DB 252 RPLITITL 260  
 |||||  
 QY 1 RPLITITL 9

Search completed: Fri Sep 11 13:41:20 1998  
 Job time : 7 secs.

\*\*\*\*\*  
 MUSE  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:41:39 1998; MasPar time 4.16 Seconds

Tabular output not generated. 91.142 Million cell updates/sec

Title: >US-08-452-843-17

Description: (1-9) from US08452843.pep

Sequence: 64

Scoring table: PAM 150

Gap 15

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database:

splemb16  
 1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
 5:sp\_mhc 6:sp\_organelle 7:sp\_phage 8:sp\_plant  
 9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp\_vertebrate  
 13:sp\_unclassified

Statistics: Mean 23.401; Variance 27.354; scale 0.855

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID  | Description            | Pred. No. |
|------------|-------|-------------|--------|--------|------------------------|-----------|
| 1          | 64    | 100.0       | 37 10  | Q64447 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |
| 2          | 64    | 100.0       | 42 4   | Q29446 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |
| 3          | 64    | 100.0       | 135 10 | Q64451 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |
| 4          | 64    | 100.0       | 136 10 | Q60434 | TUMOR SUPPRESSOR P53 ( | 3.18e-03  |
| 5          | 64    | 100.0       | 136 10 | Q64396 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |
| 6          | 64    | 100.0       | 146 4  | Q29469 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |
| 7          | 64    | 100.0       | 196 4  | Q29484 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |
| 8          | 64    | 100.0       | 205 10 | Q35873 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |
| 9          | 64    | 100.0       | 238 11 | P89004 | P53 (FRAGMENT)         | 3.18e-03  |
| 10         | 64    | 100.0       | 281 4  | Q28475 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |
| 11         | 64    | 100.0       | 285 4  | Q95326 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |
| 12         | 64    | 100.0       | 286 11 | P89003 | P53 (FRAGMENT)         | 3.18e-03  |
| 13         | 64    | 100.0       | 286 11 | P90332 | P53 (FRAGMENT)         | 3.18e-03  |
| 14         | 64    | 100.0       | 378 11 | P89002 | P53 (FRAGMENT)         | 3.18e-03  |
| 15         | 64    | 100.0       | 391 13 | Q36006 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |
| 16         | 64    | 100.0       | 393 2  | Q15086 | P53 TRANSFORMATION SUP | 3.18e-03  |
| 17         | 64    | 100.0       | 393 2  | Q15086 | P53 TRANSFORMATION SUP | 3.18e-03  |
| 18         | 64    | 100.0       | 393 2  | Q15088 | P53 TRANSFORMATION SUP | 3.18e-03  |
| 19         | 64    | 100.0       | 393 2  | Q16809 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |
| 20         | 64    | 100.0       | 393 2  | Q16811 | CELLULAR TUMOR ANTIGEN | 3.18e-03  |

|    |    |       |        |        |                        |          |
|----|----|-------|--------|--------|------------------------|----------|
| 21 | 64 | 100.0 | 393 2  | Q16808 | CELLULAR TUMOR ANTIGEN | 3.18e-03 |
| 22 | 64 | 100.0 | 393 2  | Q15087 | P53 TRANSFORMATION SUP | 3.18e-03 |
| 23 | 64 | 100.0 | 393 2  | Q16848 | CELLULAR TUMOR ANTIGEN | 3.18e-03 |
| 24 | 64 | 100.0 | 393 2  | Q16807 | CELLULAR TUMOR ANTIGEN | 3.18e-03 |
| 25 | 59 | 92.2  | 45 12  | Q92042 | CELLULAR TUMOR ANTIGEN | 5.08e-02 |
| 26 | 59 | 92.2  | 65 12  | Q92143 | CELLULAR TUMOR ANTIGEN | 5.08e-02 |
| 27 | 59 | 92.2  | 499 2  | Q15351 | P73 PROTEIN            | 5.08e-02 |
| 28 | 59 | 92.2  | 636 2  | Q15350 | P53-LIKE TRANSCRIPTION | 5.08e-02 |
| 29 | 57 | 88.1  | 634 10 | Q35834 | KET PROTEIN (FRAGMENT) | 1.49e-01 |
| 30 | 53 | 82.8  | 393 2  | Q16810 | CELLULAR TUMOR ANTIGEN | 1.20e+00 |
| 31 | 52 | 81.3  | 347 6  | Q21198 | CELLULAR TUMOR ANTIGEN | 1.98e+00 |
| 32 | 51 | 79.7  | 163 10 | Q61262 | HYPOTHETICAL 17.2 KD P | 5.34e+00 |
| 33 | 50 | 78.1  | 340 3  | Q20205 | F40810.6 (FRAGMENT)    | 5.34e+00 |
| 34 | 49 | 76.6  | 568 11 | Q69321 | UT6H.                  | 8.68e+00 |
| 35 | 48 | 75.0  | 160 9  | Q03542 | TRH PROTEIN            | 1.40e+01 |
| 36 | 48 | 75.0  | 253 3  | P94382 | DNA FOR 25-36 DEGREE R | 1.40e+01 |
| 37 | 48 | 75.0  | 1797 3 | Q93692 | F36H2.3B.              | 1.40e+01 |
| 38 | 48 | 75.0  | 1805 3 | Q93691 | F36H2.3A.              | 1.40e+01 |
| 39 | 47 | 73.4  | 347 6  | Q21143 | NADH DEHYDROGENASE SUB | 2.24e+01 |
| 40 | 47 | 73.4  | 347 6  | Q21211 | NADH DEHYDROGENASE SUB | 2.24e+01 |
| 41 | 47 | 73.4  | 347 6  | Q21675 | NADH DEHYDROGENASE SUB | 2.24e+01 |
| 42 | 47 | 73.4  | 391 9  | Q26941 | CONSERVED PROTEIN.     | 2.24e+01 |
| 43 | 47 | 73.4  | 459 9  | Q25796 | CONSERVED HYPOTHETICAL | 2.24e+01 |
| 44 | 46 | 71.9  | 237 9  | Q05330 | FOF1 ATP SYNTHASE, SUB | 3.57e+01 |
| 45 | 46 | 71.9  | 698 1  | Q04563 | HYPOTHETICAL 80.5 KD P | 3.57e+01 |

## ALIGNMENTS

RESULT 1  
 ID Q64447 PRELIMINARY; PRT; 37 AA.  
 AC Q64447;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 OS MAMMOTA MONAX (WOODCHUCK).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUDARCTIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER.  
 RA RIVKINA M.B., CULLEN J.M., ROBINSON W.S., MARION P.L.;  
 RL CANCER RES. 54:5430-5437(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER.  
 RA RIVKINA M.B., TENNANT B.C., ROBINSON W.S., MARION P.L.;  
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER.  
 RA MARION P.L.;  
 RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL; U44835; G1174225; -.  
 DR PROSITE; PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT NON\_TER 1  
 FT NON\_TER 37  
 SQ SEQUENCE 37 AA; 4140 MW; 1EBD29B4 CRC32;  
 Query Match 100.0%; Score 64; DB 10; Length 37;  
 Best Local Similarity 100.0%; Pred. No. 3.18e-03; Mismatches 0; Gaps 0;  
 Matches 9; Conservative 0; Indels 0;

DB 26 RPLITITL 34  
 1111111111  
 QY 1 RPLITITL 9

RESULT 2  
 ID 029446 PRELIMINARY; PRT; 42 AA.

AC 029446;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN TP53.

OS BOS TAURUS (BOVINE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; ARTIODACTYLA.

RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 96081357.  
 RA COGGINS L.W., SCOBIE L., JACKSON M.E., CAMPO M.S.;  
 RL NMM. GENOME 6:687-688(1995).

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; U22145; G885983; -  
 DR PROSITE; PS00348; P53; 1.  
 KM ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION.

FT NON\_TER 1  
 SQ SEQUENCE 42 AA; 4798 MW; FE01F96E CRC32;  
 Query Match 100.0%; Score: 64; DB 4; Length 42;  
 Best Local Similarity 100.0%; Pred. No. 3.18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 30 RPLITITL 38  
 1111111111  
 QY 1 RPLITITL 9

RESULT 3  
 ID 064451 PRELIMINARY; PRT; 135 AA.

AC 064451;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).

GN P53.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.

RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CD-1; TISSUE-PERT. IMPLANTATION BLASTOCYSTS;  
 RA LALONAVA M., KOMAR G.P., KOIDE S.S.;  
 RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDJ DATA BANKS.

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; U59757; G1399878; -  
 DR PROSITE; PS00348; P53; 1.  
 KM ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT NON\_TER 1

FT NON\_TER 135 135  
 SQ SEQUENCE 135 AA; 15317 MW; 3C3655B9 CRC32;

Query Match 100.0%; Score: 64; DB 10; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 3.18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 125 RPLITITL 133  
 1111111111  
 QY 1 RPLITITL 9

RESULT 4  
 ID 060434 PRELIMINARY; PRT; 136 AA.

AC 060434; P97257;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE TUMOR SUPPRESSOR P53 (FRAGMENT).

OS CRICETULUS GRISEUS (CHINESE HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.

RN 11  
 RP SEQUENCE FROM N.A.  
 RA MAI S., FLURI M., SIWASKI D., HUPPI K.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDJ DATA BANKS.

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; U41452; G1786175; -  
 DR PROSITE; PS00348; P53; 1.  
 KM ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION.

FT NON\_TER 1  
 SQ SEQUENCE 136 AA; 15438 MW; 10679AD4 CRC32;  
 Query Match 100.0%; Score: 64; DB 10; Length 136;  
 Best Local Similarity 100.0%; Pred. No. 3.18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 118 RPLITITL 126  
 1111111111  
 QY 1 RPLITITL 9

RESULT 5  
 ID 064396 PRELIMINARY; PRT; 136 AA.

AC 064396; P97940;  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).

GN P53.  
 OS CRICETULUS GRISEUS (CHINESE HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.

RN 11  
 RP SEQUENCE FROM N.A.  
 RA MAI S., FLURI M., SIWASKI D., HUPPI K.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDJ DATA BANKS.

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; U41451; G1786175; -



DR PROSITE; PS00348; P53; 1.  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ANTI-ONCOGENE; DNA-BINDING;  
 KW TRANSCRIPTION REGULATION; ACTIVATOR.  
 FT NON\_TER 136 136  
 FT NON\_TER 136 136  
 SQ SEQUENCE 136 AA; 15411 MW; CFB916C9 CRC32;  
 Query Match 100.0%; Score 64; DB 10; Length 136;  
 Best Local Similarity 100.0%; Pred. No. 3,18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 118 RPLTITL 126  
 OY 1 RPLTITL 9  
 RESULT 6  
 ID Q29469 PRELIMINARY; PRT; 146 AA.  
 AC Q29469;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS CANIS FAMILIARIS (DOG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; CARNIVORA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA NARSIR L., MCFARLANE S.T., ARGYLE D.J., REID S.W.J.;  
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; U51857; G1263295; -.  
 DR PROSITE; PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT NON\_TER 146 146  
 FT NON\_TER 146 146  
 SQ SEQUENCE 146 AA; 16396 MW; 8AE726C9 CRC32;  
 Query Match 100.0%; Score 64; DB 4; Length 146;  
 Best Local Similarity 100.0%; Pred. No. 3,18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 115 RPLTITL 123  
 OY 1 RPLTITL 9  
 RESULT 7  
 ID Q29484 PRELIMINARY; PRT; 196 AA.  
 AC Q29484;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS EQUUS CABALLUS (HORSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PERISSODACTYLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BUCHER K., SZALAI G., MARTI E., PAULI U., LAZARY S.;  
 RL RES. VET. SCI. 0:0-0(0).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY

CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; X91793; E218035; -.  
 DR PROSITE; PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT NON\_TER 196 196  
 FT NON\_TER 196 196  
 SQ SEQUENCE 196 AA; 22080 MW; F443239C CRC32;  
 Query Match 100.0%; Score 64; DB 10; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 3,18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 116 RPLTITL 124  
 OY 1 RPLTITL 9  
 RESULT 8  
 ID Q35873 PRELIMINARY; PRT; 205 AA.  
 AC Q35873;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS CRICETUS GRISEUS (CHINESE HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA RAINALDI G., MARCETTI S., CAPECCHI B., MENEVERI R., PIRAS A.,  
 RA LEUZZI R.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA VATERONI L., MUSIO A., MENEVERI R., RAINALDI G.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; U74487; G2581764; -.  
 DR PROSITE; PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT NON\_TER 205 205  
 FT NON\_TER 205 205  
 SQ SEQUENCE 205 AA; 23122 MW; 680DDDC CRC32;  
 Query Match 100.0%; Score 64; DB 10; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 3,18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 124 RPLTITL 132  
 OY 1 RPLTITL 9  
 RESULT 9  
 ID P89004 PRELIMINARY; PRT; 238 AA.  
 AC P89004;  
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)  
 DE P53 (FRAGMENT).  
 OS MASTOMYS NATALENSIS PAPILLOMAVIRUS (MNPV).

OC VIRIDAE: DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.  
 RN [1] SEQUENCE FROM N.A.  
 RC TISSUE-ECLONIA INDUCED BY LOXTIDINE;  
 RA LUQUE E.A., TANG L.H., MODLIN I.M.;  
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U48618; G1813455; -  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 FT NON\_TER 1  
 SQ SEQUENCE 238 AA; 26704 MW; 097E01F9 CRC32;

Query Match  
 Best Local Similarity 100.0%; Score 64; DB 11; Length 238;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 142 RPLTITL 150  
 OY 1 RPLTITL 9

RESULT 10  
 ID 029475 PRELIMINARY; PRT; 281 AA.  
 AC 029475;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 CC P53  
 GN CANIS FAMILIARIS (DOG).  
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; CARNIVORA.  
 RN [1]  
 RP TISSUE-MAMMARY GLAND;  
 RC MEDLINE: 97194812;  
 RA VAN LEEUWEN I., RUTEMAN G.R., HELLMEN E., CORNELISSE C.C.J.,  
 RA DEVILIER P.;  
 RL ANTICANCER RES. 16:3737-3744(1996).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC -1- PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: L37107; G1463021; -  
 DR PROSITE: P500348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT NON\_TER 1  
 SQ SEQUENCE 281 AA; 31762 MW; FCTBAE31 CRC32;

Query Match  
 Best Local Similarity 100.0%; Score 64; DB 4; Length 281;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 142 RPLTITL 150  
 OY 1 RPLTITL 9

RESULT 11  
 ID 095326 PRELIMINARY; PRT; 285 AA.  
 AC 095326;  
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
 DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 CC P53  
 GN CANIS FAMILIARIS (DOG).  
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; CARNIVORA.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA YANG B.J., SHI X.B., LAU D.H.M.;  
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: U62133; G1619833; -  
 DR PROSITE: P500348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT NON\_TER 1  
 SQ SEQUENCE 285 AA; 31616 MW; 15E1EC47 CRC32;

Query Match  
 Best Local Similarity 100.0%; Score 64; DB 4; Length 285;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 204 RPLTITL 212  
 OY 1 RPLTITL 9

RESULT 12  
 ID P89003 PRELIMINARY; PRT; 286 AA.  
 AC P89003;  
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)  
 DE P53 (FRAGMENT).  
 CC MASTOMYS NATALENSIS PAPILLOMAVIRUS (MNPV).  
 CC VIRIDAE: DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA LUQUE E.A., TANG L.H., MODLIN I.M.;  
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U48617; G1813453; -  
 FT NON\_TER 1  
 SQ SEQUENCE 286 AA; 32287 MW; 30F7C9FA CRC32;

Query Match  
 Best Local Similarity 100.0%; Score 64; DB 11; Length 286;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 142 RPLTITL 150  
 OY 1 RPLTITL 9

RESULT 13  
 ID P90332 PRELIMINARY; PRT; 286 AA.  
 AC P90332;  
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)  
 DE P53 (FRAGMENT).  
 CC MASTOMYS NATALENSIS PAPILLOMAVIRUS (MNPV).  
 CC VIRIDAE: DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-SPONTANEOUS ECLONAS;  
 RA LUQUE E.A., TANG L.H., MODLIN I.M.;  
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U48619; G1813457; -  
 FT NON\_TER 1  
 SQ SEQUENCE 286 AA; 32247 MW; 5B5D3CAD CRC32;

Query Match  
 Best Local Similarity 100.0%; Score 64; DB 11; Length 286;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB: 142 RPLTITL 150  
 |||||  
 OY 1 RPLTITL 9

## RESULT 14

PRELIMINARY: PRT: 378 AA.

ID P89002  
 AC P89002;  
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)  
 DE P53 (FRAGMENT).  
 OS MASTOMYS NATALENSIS PAPILLOMAVIRUS (MNPV).  
 OC VIRIDAE: D6-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA LUDUE E.A., TANG L.H., MODLIN I.M.;  
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U48616; G1813451;  
 FT NON\_TER 1  
 SO SEQUENCE 378 AA; 42062 MW; B4436760 CRC32;

## Query Match

100.0%; Score 64; DB 11; Length 378;

Best Local Similarity 100.0%; Pred. No. 3.18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 234 RPLTITL 242  
 |||||  
 OY 1 RPLTITL 9

## RESULT 15

PRELIMINARY: PRT: 391 AA.

ID 036006  
 AC 036006;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN P53.  
 OS MARMATA MONAX.  
 OG PLASMID PTBLUE (R).  
 OC UNCLASSIFIED.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA FEITELSON M.A., RANGANATHAN P.N., CLAYTON M.M., ZHANG S.M.;  
 RL ONCOGENE 15:327-336(1997).  
 -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: AJ001022; E351287;  
 DR PROSITE: PS00346; P53; 1.  
 DR PLASMID: ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION;  
 KW ACTIVATOR; NUCLEAR PROTEIN; PHOSPHORYLATION.  
 SO SEQUENCE 391 AA; 43468 MW; 95FAB8F2 CRC32;

## Query Match

100.0%; Score 64; DB 13; Length 391;

Best Local Similarity 100.0%; Pred. No. 3.18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 247 RPLTITL 255  
 |||||  
 OY 1 RPLTITL 9

Search completed: Fri Sep 11 13:42:12 1998  
 Job time : 33 secs.

**THIS PAGE BLANK (USPTO)**

# WIRE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:36:12 1998; Maspar time 2.70 Seconds  
53.996 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-16  
Description: (1-9) from US08452843.pep  
Perfect Score: 63  
Sequence: 1 GTRVRAMAI 9

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: 1  
a.geneseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 15.795; Variance 44.115; scale 0.358

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length DB | ID     | Description           | Pred. No. |
|------------|-------|-------------|-----------|--------|-----------------------|-----------|
| 1          | 63    | 100.0       | 241 10    | R51872 | Human p53 amino acids | 7.75e-01  |
| 2          | 63    | 100.0       | 253 24    | W28484 | Human p53 protein var | 7.75e-01  |
| 3          | 63    | 100.0       | 253 24    | W28483 | Human p53 protein var | 7.75e-01  |
| 4          | 63    | 100.0       | 270 24    | W28486 | Human p53 protein var | 7.75e-01  |
| 5          | 63    | 100.0       | 270 24    | W28485 | Human p53 protein var | 7.75e-01  |
| 6          | 63    | 100.0       | 319 24    | W28495 | Human p53 protein var | 7.75e-01  |
| 7          | 63    | 100.0       | 319 24    | W28496 | Human p53 protein var | 7.75e-01  |
| 8          | 63    | 100.0       | 328 10    | R51876 | Human p53 amino acids | 7.75e-01  |
| 9          | 63    | 100.0       | 325 24    | W28497 | Human p53 protein var | 7.75e-01  |
| 10         | 63    | 100.0       | 325 24    | W28498 | Human p53 protein var | 7.75e-01  |
| 11         | 63    | 100.0       | 353 24    | W28494 | Human p53 protein var | 7.75e-01  |
| 12         | 63    | 100.0       | 353 24    | W28493 | Human p53 protein var | 7.75e-01  |
| 13         | 63    | 100.0       | 361 21    | W13958 | Chimeric p53 protein. | 7.75e-01  |
| 14         | 63    | 100.0       | 361 21    | W13958 | Chimeric p53 protein. | 7.75e-01  |
| 15         | 63    | 100.0       | 363 21    | W13973 | Modified p53 variant  | 7.75e-01  |
| 16         | 63    | 100.0       | 363 21    | W13972 | Modified p53 variant  | 7.75e-01  |
| 17         | 63    | 100.0       | 363 24    | W28479 | Human p53 protein var | 7.75e-01  |
| 18         | 63    | 100.0       | 363 24    | W28480 | Human p53 protein var | 7.75e-01  |

|    |    |       |        |        |                       |          |
|----|----|-------|--------|--------|-----------------------|----------|
| 19 | 63 | 100.0 | 363 21 | W13976 | Modified p53 variant  | 7.75e-01 |
| 20 | 63 | 100.0 | 368 21 | W13956 | Chimeric p53 protein. | 7.75e-01 |
| 21 | 63 | 100.0 | 370 21 | W13957 | Chimeric p53 protein. | 7.75e-01 |
| 22 | 63 | 100.0 | 374 24 | W28481 | Human p53 protein var | 7.75e-01 |
| 23 | 63 | 100.0 | 374 24 | W28482 | Human p53 protein var | 7.75e-01 |
| 24 | 63 | 100.0 | 381 24 | W28490 | Human p53 protein var | 7.75e-01 |
| 25 | 63 | 100.0 | 381 24 | W28489 | Human p53 protein var | 7.75e-01 |
| 26 | 63 | 100.0 | 393 24 | W25155 | Human p53 variant fou | 7.75e-01 |
| 27 | 63 | 100.0 | 393 22 | W13948 | Human wild-type p53 t | 7.75e-01 |
| 28 | 63 | 100.0 | 393 21 | W05345 | Human mutant N239     | 7.75e-01 |
| 29 | 63 | 100.0 | 393 22 | W13951 | Human tumour-derivate | 7.75e-01 |
| 30 | 63 | 100.0 | 393 21 | W05347 | Human p53 mutant R248 | 7.75e-01 |
| 31 | 63 | 100.0 | 393 21 | W05346 | Human p53 mutant R273 | 7.75e-01 |
| 32 | 63 | 100.0 | 393 19 | W02517 | Human p53 tumour supp | 7.75e-01 |
| 33 | 63 | 100.0 | 393 18 | R91833 | Wild type p53 protein | 7.75e-01 |
| 34 | 63 | 100.0 | 393 21 | W13968 | Modified p53 variant  | 7.75e-01 |
| 35 | 63 | 100.0 | 393 21 | W13970 | Modified p53 variant  | 7.75e-01 |
| 36 | 63 | 100.0 | 393 21 | W13969 | Modified p53 variant  | 7.75e-01 |
| 37 | 63 | 100.0 | 401 24 | W28487 | Human p53 protein var | 7.75e-01 |
| 38 | 63 | 100.0 | 401 24 | W28488 | Human p53 protein var | 7.75e-01 |
| 39 | 63 | 100.0 | 406 21 | W13966 | Chimeric p53 protein. | 7.75e-01 |
| 40 | 63 | 100.0 | 406 21 | W13964 | Chimeric p53 protein. | 7.75e-01 |
| 41 | 63 | 100.0 | 411 21 | W13967 | Chimeric p53 protein. | 7.75e-01 |
| 42 | 63 | 100.0 | 438 14 | R74272 | Tumour suppressor pro | 7.75e-01 |
| 43 | 63 | 100.0 | 533 23 | W19763 | p53-GM-CSF immunostim | 7.75e-01 |
| 44 | 63 | 100.0 | 535 24 | W28492 | Human p53 protein var | 7.75e-01 |
| 45 | 63 | 100.0 | 535 24 | W28491 | Human p53 protein var | 7.75e-01 |

## ALIGNMENTS

RESULT 1  
ID R51872 standard; Protein; 241 AA.  
AC R51872;  
DE 18-NOV-1994 (first entry)  
DT Human p53 amino acids 1-241.  
KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
KW anti-oncogene; cancer; tumour; antibody binding region; epitope.  
OS Homo sapiens.  
PN W09408241-A.  
PD 14-APR-1994.  
PE 30-SEP-1993; E02666.  
PF 30-SEP-1993; DE-232823.  
PR 30-SEP-1993; DE-232823.  
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
PI Klein R, Schanz P, Tessmer C, Volkmann M, Zentgraf H;  
DR N-PSDB; 062357.  
DR Non-radioactive detection of p53 specific antibodies - by capture  
PT on immobilised p53 or its fragments, then reaction with labelled  
PT second antibody, for diagnosis of tumours and suitable for  
PT screening  
PS Claim 10; Page 17; 35pp; German.  
CC Antibodies specific for p53 are detected by binding to immobilised  
CC fragments of the p53 gene product containing the antibody-binding  
CC region. Preferred fragments contain amino acids 1-241, 40-349,  
CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
CC 368-386. See R51872-R51881 for sequences of these fragments.  
SQ Sequence 241 AA;

Query Match 100.0%; Score 63; DB 10; Length 241;  
Best Local Similarity 100.0%; Pred. No. 7.75e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 154 gtrvramai 162  
|||  
OY 1 GTRVRAMAI 9  
RESULT 2  
ID W28484 standard; Protein; 253 AA.  
AC W28484;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant V-367H.

KM Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KM anti-oncogene; hyperproliferation; cancer; restenosis;  
 OS tumour suppression; apoptosis;  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_difference 189  
 FT /note- "Arg residue at position 182 of wild-type  
 p53 has been mutated to His"  
 EN MO9704092-A1  
 PD 06-FEB-1997  
 PR 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Consellier E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 leucine zipper - useful for treating hyper-proliferative disorders,  
 esp. cancer and restenosis  
 PS Claim 32; Page -: 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the p53  
 transactivation domain (amino acids 1-74) deleted and replaced by  
 the transactivating domain (TD) from herpes simplex virus viral  
 protein VP16 (amino acids 411-490). The present sequence is that of  
 a specifically claimed p53 variant designated V-367 and comprising  
 the VP16 TD and amino acids 75-367 of human wild-type p53 (but with  
 Arg182 replaced by His). The p53 variants are more active and more  
 stable tumour suppressors and apoptosis-inducing agents than wild-type  
 p53 and are active where the wild-type protein is not.  
 CC (Note: this sequence does not appear in the specification and has  
 been produced by modifying the given sequence of variant V-367).  
 CC Sequence 253 AA;  
 SQ

Query Match 100.0%; Score 63; DB 24; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 7.75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 162 gtrvramai 170  
 ||||||||  
 1 GTRVRAMAI 9

QY

RESULT 3  
 ID W28483 standard; Protein: 253 AA.  
 AC W28483;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-367 encoded by PEC141.  
 KM Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KM anti-oncogene; hyperproliferation; cancer; restenosis;  
 OS tumour suppression; apoptosis;  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 PN MO9704092-A1.  
 PD 06-FEB-1997.  
 PR 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Consellier E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 leucine zipper - useful for treating hyper-proliferative disorders,  
 esp. cancer and restenosis  
 PS Claim 32; Pages 80-81; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the p53  
 transactivation domain (amino acids 1-74) deleted and replaced by  
 the transactivating domain (TD) from herpes simplex virus viral  
 protein VP16 (amino acids 411-490). The present sequence is that of  
 a specifically claimed p53 variant designated V-367 and comprising

CC the VP16 TD with amino acids 75-367 of human wild-type p53. The p53  
 CC variants are more active and more stable tumour suppressors and  
 CC apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not.  
 SQ Sequence 253 AA;  
 SQ

Query Match 100.0%; Score 63; DB 24; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 7.75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 162 gtrvramai 170  
 ||||||||  
 1 GTRVRAMAI 9

QY

RESULT 4  
 ID W28486 standard; Protein: 270 AA.  
 AC W28486;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-ASh.  
 KM Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KM anti-oncogene; hyperproliferation; cancer; restenosis; murine;  
 KM tumour suppression; apoptosis; alternative splicing; AS form.  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_difference 189  
 FT /note- "Arg residue at position 182 of wild-type  
 p53 has been mutated to His"  
 EN MO9704092-A1.  
 PD 06-FEB-1997.  
 PR 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Consellier E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 leucine zipper - useful for treating hyper-proliferative disorders,  
 esp. cancer and restenosis  
 PS Claim 33; Page -: 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the p53  
 transactivation domain (amino acids 1-74) deleted and replaced by  
 the transactivating domain (TD) from herpes simplex virus viral  
 protein VP16 (amino acids 411-490). The present sequence is that of  
 a specifically claimed p53 variant designated V-ASh and comprising  
 the VP16 TD with amino acids 75-366 of human wild-type p53 (but  
 CC with Arg182 replaced by His), followed by the last 19 C-terminal  
 CC amino acids of the alternatively spliced (AS) form of murine p53  
 CC (encoded by a synthetic linker). The variants are more active and  
 CC more stable tumour suppressors and apoptosis-inducing agents than  
 CC wild-type p53 and are active where the wild-type protein is not.  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant V-AS).  
 CC Sequence 270 AA;  
 SQ

Query Match 100.0%; Score 63; DB 24; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 7.75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 162 gtrvramai 170  
 ||||||||  
 1 GTRVRAMAI 9

QY

RESULT 5  
 ID W28485 standard; Protein: 270 AA.  
 AC W28485;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-AS encoded by PEC143.  
 KM Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;

KW anti-oncogene; hyperproliferation; cancer; restenosis; murine;  
 KM tumour suppression; apoptosis; alternative splicing; AS form.  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 PD WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB; T86218.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 leucine zipper - useful for treating hyper-proliferative disorders,  
 esp. cancer and restenosis  
 PS Claim 33: Pages 62-83; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the p53  
 transactivation domain (amino acids 1-74) deleted and replaced by  
 the transactivating domain (TD) from herpes simplex virus viral  
 protein VP16 (amino acids 411-490). The present sequence is that of  
 a specifically claimed p53 variant designated V-AS and comprising  
 the VP16 TD with amino acids 75-366 of human wild-type p53, followed  
 by the last 19 C-terminal amino acids of the alternatively spliced  
 (AS) form of murine p53 (encoded by a synthetic linker). The  
 CC variants are more active and more stable tumour suppressors and  
 CC apoptosis-inducing agents than wild-type p53 and are active where  
 the wild-type protein is not.  
 SO Sequence 270 AA;

Query Match 100.0%; Score 63; DB 24; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 7,75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 162 gtrvramal 170  
 ||||||||  
 QY 1 GTRVRAMAI 9

RESULT 6  
 ID W28495 standard; Protein: 319 AA.  
 AC W28495;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360-325 encoded by p53178.  
 KM Leucine zipper domain; LZD; oligomerisation domain; mutant; mutetin;  
 KM substitution; replacement; transactivation; viral protein VP16; HSV;  
 KM anti-oncogene; hyperproliferation; cancer; restenosis;  
 KM tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 PD WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB; T86223.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 leucine zipper - useful for treating hyper-proliferative disorders,  
 esp. cancer and restenosis  
 PS Claim 38: Pages 92-94; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 oligomerisation domain deleted and replaced by a leucine zipper  
 domain. The mutants preferably also have at least part of the p53  
 transactivation domain (amino acids 1-74) deleted and replaced by  
 the domain 325-360 of p53. The present sequence is that of a  
 specifically claimed p53 variant designated 360-325 and comprising  
 the 325-360 domain, amino acids 75-325 of human wild-type p53 and a  
 leucine zipper domain at the C-terminal. The p53 variants are  
 more active and more stable tumour suppressors and apoptosis-inducing  
 agents than wild-type p53 and are active where the wild-type protein  
 is not, i.e. they are not inactivated by dominant negative or oncogenic

CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SO Sequence 319 AA;

Query Match 100.0%; Score 63; DB 24; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 7,75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 118 gtrvramal 126  
 ||||||||  
 QY 1 GTRVRAMAI 9

RESULT 7  
 ID W28496 standard; Protein: 319 AA.  
 AC W28496;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360-325H.  
 KM Leucine zipper domain; LZD; oligomerisation domain; mutant; mutetin;  
 KM substitution; replacement; transactivation; viral protein VP16; HSV;  
 KM anti-oncogene; hyperproliferation; cancer; restenosis;  
 KM tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 FH Key Location/qualifiers  
 FT misc\_difference 145  
 FT /note- "Arg residue at position 182 of wild-type  
 p53 has been mutated to His"  
 FT  
 PD WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 leucine zipper - useful for treating hyper-proliferative disorders,  
 esp. cancer and restenosis  
 PS Claim 38: Page -; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 oligomerisation domain deleted and replaced by a leucine zipper  
 domain. The mutants preferably also have at least part of the p53  
 transactivation domain (amino acids 1-74) deleted and replaced by  
 the domain 325-360 of p53. The present sequence is that of a  
 specifically claimed p53 variant designated 360-325H and comprising  
 the 325-360 domain, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 360-325).  
 SO Sequence 319 AA;

Query Match 100.0%; Score 63; DB 24; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 7,75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 118 gtrvramal 126  
 ||||||||  
 QY 1 GTRVRAMAI 9

RESULT 8  
 ID R51876 standard; Protein: 328 AA.  
 AC R51876;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 66-393.  
 KM Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 KM anti-oncogene; cancer; tumour; antibody binding region; epitope.

OS Homo sapiens. Location/Qualifiers  
 FH Key misc\_difference 208 /note= "Arg corresponds to a CAT codon"  
 FT  
 FT  
 PN W090408241-A.  
 PD 14-APR-1994.  
 PE 30-SEP-1993; E02666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schranz P, Tessmer C, Volkmann M, Zentgraf H;  
 DR WPI: 94.135732/16.  
 DT N-PSDB: 062324.  
 PT Non-radioactive detection of p53 specific antibodies - by capture  
 PT on immobilised p53 or its fragments, then reaction with labelled  
 PT second antibody for diagnosis of tumours and suitable for  
 PT screening  
 PS Claim 10: Page 18; 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 SQ Sequence 328 AA;  
 Query Match 100.0%; Score 63; DB 10; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 7.75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 89 gtrvramai 97  
 |||||  
 QY 1 GTRVRAMAI 9  
 RESULT 9  
 ID W28497 standard; Protein: 335 AA.  
 AC W28497;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360h-325 encoded by PEC179.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muteln;  
 KW substitution; replacement; transactivation; hinge region;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT region 39..53  
 FT /label= hinge  
 PN W09704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON) RHONE-POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DT N-PSDB: T86224.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 39: Pages 94-95; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360h-325 and comprising  
 CC the 325-360 domain, separated from amino acids 75-325 of human  
 CC wild-type p53 by a synthetic hinge sequence (Gly4Ser)3, and with a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).

SQ Sequence 335 AA;  
 Query Match 100.0%; Score 63; DB 24; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 7.75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 134 gtrvramai 142  
 |||||  
 QY 1 GTRVRAMAI 9  
 RESULT 10  
 ID W28498 standard; Protein: 335 AA.  
 AC W28498;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360h-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muteln;  
 KW substitution; replacement; transactivation; hinge region;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT region 39..53  
 FT /label= hinge  
 FT misc\_difference 161 /note= "Arg residue at position 182 of wild-type  
 p53 has been mutated to His"  
 PN W09704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON) RHONE-POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DT New p53 variants e.g. with oligomerisation domain replaced by  
 DT leucine zipper - useful for treating hyper-proliferative disorders,  
 DT esp. cancer and restenosis  
 PS Claim 39: Page -; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360h-325H and comprising  
 CC the 325-360 domain, separated from amino acids 75-325 of human  
 CC wild-type p53 (but with Arg182 replaced by His) by a synthetic hinge  
 CC sequence (Gly4Ser)3, and with a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 360h-325).  
 SQ Sequence 335 AA;  
 Query Match 100.0%; Score 63; DB 24; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 7.75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 134 gtrvramai 142  
 |||||  
 QY 1 GTRVRAMAI 9  
 RESULT 11  
 ID W28494 standard; Protein: 353 AA.  
 AC W28494;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 393-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muteln;



KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KM anti-oncogene; hyperproliferation; cancer; restenosis;  
 KM tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 FH Key  
 FT msc.difference 179  
 FT /note="Arg residue at position 182 of wild-type  
 FT p53 has been mutated to His"  
 PD MO9704092-A1.  
 PD 06-FEB-1997.  
 PE 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 37; Page -; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-393 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 393-325H and comprising  
 CC the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 393-325).  
 SQ Sequence 353 AA.

Query Match 100.0%; Score 63; DB 24; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 7.75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 152 gtrvram1 160.  
 QY 1 GTRVRAAI 9

RESULT 12  
 ID W28493 standard; Protein: 353 AA.  
 AC W28493;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 393-325 encoded by PEC177.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; Viral protein VP16; HSV;  
 KM anti-oncogene; hyperproliferation; cancer; restenosis;  
 KM tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PE 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 37; Pages 90-92; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53

CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-393 of p53. The present sequence is that of  
 CC a specifically claimed p53 variant designated 393-325 and comprising  
 CC the 325-393 domain, amino acids 75-325 of human wild-type p53 and a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 353 AA.

Query Match 100.0%; Score 63; DB 24; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 7.75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 152 gtrvram1 160  
 QY 1 GTRVRAAI 9

RESULT 13  
 ID W13961 standard; Protein: 361 AA.  
 AC W13961;  
 DT 25-JUN-1997 (first entry)  
 DE Chimeric p53 protein.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; GCN4; DNA binding.  
 OS Chimeric Homo sapiens;  
 OS Chimeric synthetic.  
 FH Key  
 FT Location/Qualifiers  
 FT region 1..323  
 FT /label= p53wt  
 FT /note="amino acids 1-323 of wild-type p53"  
 FT region 324..329  
 FT /label= Linker  
 FT region 330..361  
 FT /label= GCN4  
 FT /note="amino acids 250-281 of GCN4 LZ variant"  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PE 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazometis TD;  
 DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Disclosure: Refer to Page 8; 82pp; English.  
 CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
 CC of human wild-type p53 tumour suppressor (see also W13948) linked  
 CC to a C-terminal portion of the LZ variant (see also W13955) of  
 CC GCN4 and, in some cases, the C-terminal portion of wild-type  
 CC p53. The chimeric proteins have DNA binding activity and can  
 CC replace lost or insufficient p53 function, providing the means for  
 CC pharmacological rescue of p53 function in cancer patients. Nucleic  
 CC acids coding for modified p53 constructs can be used for cancer  
 CC gene therapy.  
 SQ Sequence 361 AA.

Query Match 100.0%; Score 63; DB 21; Length 361;  
 Best Local Similarity 100.0%; Pred. No. 7.75e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 154 gtrvram1 162  
 QY 1 GTRVRAAI 9

RESULT 14  
 ID W13958 standard; Protein: 361 AA.  
 AC W13958;



(W.I.)

Distribution rights by Oxford Molecular Ltd

un - protein database search

Fri Sep 11 13:36:39 1998; MasPar time 3.08 seconds

generated.

(1-9) from US08452843.pdf

1 GTRVRAMAI 9

2AM 1.50

120441 seqs, 36531193 residues

existing first 45 summaries

01r56

Mean 22.295; Variance 28.032; scale 0.795

erved by analysis of the total score distribution.

| Entry | Length | DB | ID      | Description            | Pred. No |
|-------|--------|----|---------|------------------------|----------|
| 0.0   | 191    | 5  | 1CSA    | p53 residues 97-287,   | 5,188-00 |
| 0.0   | 194    | 5  | 1IUPB   | tumor suppressor p53,  | 5,188-00 |
| 0.0   | 194    | 5  | 1ISRB   | p53 tumor suppressor,  | 5,188-00 |
| 0.0   | 195    | 5  | 1ISRC   | tumor suppressor p53,  | 5,188-00 |
| 0.0   | 195    | 5  | 1ISRC   | p53 tumor suppressor,  | 5,188-00 |
| 0.0   | 196    | 5  | 1IUPA   | tumor suppressor p53,  | 5,188-00 |
| 0.0   | 196    | 5  | 1ISRA   | p53 tumor suppressor,  | 5,188-00 |
| 0.0   | 386    | 2  | SS168   | cellular tumor antigen | 5,188-00 |
| 0.0   | 391    | 2  | UC6193  | tumor suppressor p53   | 5,188-00 |
| 0.0   | 391    | 2  | SO2192  | cellular tumor antigen | 5,188-00 |
| 0.0   | 393    | 2  | UC6116  | tumor suppressor prot  | 5,188-00 |
| 0.0   | 393    | 1  | DNHU53  | cellular tumor antigen | 5,188-00 |
| 0.0   | 396    | 2  | UD0633  | cellular tumor antigen | 5,188-00 |
| 0.0   | 381    | 2  | S388824 | cellular tumor antigen | 2,558-00 |
| 0.0   | 390    | 1  | DNM853  | cellular tumor antigen | 2,558-00 |
| 0.0   | 393    | 2  | SO6594  | cellular tumor antigen | 2,558-00 |
| 0.0   | 564    | 2  | BS9137  | sensory transduction   | 5,466-00 |
| 0.0   | 567    | 2  | HS9145  | sensory transduction   | 8,946-00 |
| 0.0   | 448    | 2  | S388658 | DNA-directed RNA poly  | 1,466-00 |
| 0.0   | 386    | 2  | UC4865  | contractile tail shea  | 3,806-00 |
| 0.0   | 387    | 1  | UC5191  | tail sheath protein    | 3,806-00 |
| 0.0   | 493    | 2  | XBC     | acetyl-CoA C-acetylran | 6,086-00 |
| 0.0   | 493    | 2  | AS9160  | sensory transduction   | 6,086-00 |

|    |    |      |      |      |        |                        |                       |         |
|----|----|------|------|------|--------|------------------------|-----------------------|---------|
| 45 | 45 | 44   | 65.8 | 1953 | 2      | 563244                 | hypoethetical protein | 1.53e+0 |
| 44 | 44 | 65.8 | 244  | 2    | D6758  | DNA ligase (ATP) (EC   | 1.53e+0               |         |
| 25 | 26 | 74.6 | 573  | 2    | D64321 | early light-inducible  | 2.40e+0               |         |
| 27 | 27 | 73.0 | 192  | 3    | J05876 | spou protein - Escher  | 2.40e+0               |         |
| 28 | 28 | 73.0 | 229  | 2    | U00043 | biphenyl dioxygenase   | 2.40e+0               |         |
| 29 | 29 | 72.0 | 408  | 1    | E42409 | biphenyl dioxygenase   | 2.40e+0               |         |
| 32 | 31 | 72.0 | 409  | 1    | D29830 | benzene 1,2-dioxygena  | 2.40e+0               |         |
| 31 | 32 | 71.4 | 410  | 1    | D6516  | toluene dioxygenase (  | 3.74e+0               |         |
| 33 | 33 | 71.4 | 189  | 1    | RKHS   | ribulose-bisphosphate  | 3.74e+0               |         |
| 33 | 45 | 71.4 | 504  | 2    | A57215 | glial cells missing (  | 3.74e+0               |         |
| 34 | 44 | 65.8 | 92   | 2    | S43106 | orfil protein - yeast  | 5.79e+0               |         |
| 35 | 35 | 65.8 | 179  | 1    | R5B5   | Ribosomal protein L5   | 5.79e+0               |         |
| 37 | 37 | 65.8 | 191  | 2    | S29884 | Ribosomal protein L5   | 5.79e+0               |         |
| 39 | 39 | 65.8 | 302  | 2    | S76323 | hypoethetical protein  | 5.79e+0               |         |
| 39 | 44 | 65.8 | 369  | 2    | S53772 | farnesyltransferase    | 5.79e+0               |         |
| 39 | 44 | 65.8 | 382  | 2    | B59877 | sulfate adenylyltrans  | 5.79e+0               |         |
| 40 | 40 | 65.8 | 443  | 2    | D59306 | conserved hypoethica   | 5.79e+0               |         |
| 41 | 41 | 65.8 | 450  | 2    | B69168 | UDP-N-acetylmuramyl    | 5.79e+0               |         |
| 42 | 42 | 65.8 | 469  | 2    | D44661 | ATP synthase F1, subu  | 5.79e+0               |         |
| 43 | 43 | 65.8 | 486  | 2    | I39523 | hydroethosulimate dehy | 5.79e+0               |         |
| 44 | 44 | 65.8 | 501  | 2    | S22669 | hypoethetical protein  | 5.79e+0               |         |
| 45 | 44 | 65.8 | 1953 | 2    | 563244 | BNI1 protein - yeast   | 5.79e+0               |         |

## ALIGNMENTS

|                       |   |  |
|-----------------------|---|--|
| RESULT                | 1   |  |
| ENTRY                 |   |  |
| TITLE                 | 1YCSA #type complete  |  |
| PDB_TITLE             | p53 residues 97-287, chain A - human                          |  |
| ORGANISM              | p53-53bp2 complex   |  |
| note                  | #formal_name Homo sapiens #common_name man                    |  |
| REFERENCE             | expressed in Escherichia coli, strain b121 (d3)               |  |
| #authors              | A68208  |  |
| #submission           | Gorina, S.; Pavletich, N.P.                                   |  |
| #cross-references     | submitted to the Brookhaven Protein Data Bank, September 1996 |  |
| REFERENCE             | #cross-references PDB:1YCS                                    |  |
| #authors              | TN001216  |  |
| #journal              | Gorina, S.; Pavletich, N.P.                                   |  |
| #title                | Science (1996) 274:1001                                       |  |
|                       | Structure of the p53 tumor suppressor bound to the ankyrin    |  |
| REFERENCE             | and sh3 domains of 53bp2.                                     |  |
| #authors              | TN001217  |  |
| #journal              | Naumovski, L.; Cleary, M.L.                                   |  |
| #title                | Mol. Cell. Biol. (1996) 16:3984                               |  |
|                       | The p53-binding protein 53bp2 also interacts with bcl2 and    |  |
| REFERENCE             | impedes cell cycle progression at g2w.                        |  |
| #authors              | I38604  |  |
|                       | Iwabuuchi, K.; Bartel, P.L.; Li, B.; Marriacino, R.; Fields,  |  |
|                       | S.  |  |
|                       | Proc. Natl. Acad. Sci. U.S.A. (1994) 91:6098-6102             |  |
|                       | Two cellular proteins that bind to wild-type but not mutant   |  |
|                       | p53.  |  |
| #cross-references     | #cross-references NUID:94286584                               |  |
| COMMENT               | Resolution: 2.2 angstroms                                     |  |
| COMMENT               | Determination: X-ray diffraction                              |  |
| COMMENT               | R-value: 0.205  |  |
| KEYWORDS              | ankyrin repeats; anti-oncogene; complex; disease mutation     |  |
|                       | polymorphism; multigene family; nuclear protein; p53;         |  |
|                       | phosphorylation; sh3; tumor suppressor                        |  |
| FEATURE               |   |  |
| 9-11                  | #region helix (right hand 3-10)\                              |  |
| 81-85                 | #region helix (right hand alpha)\                             |  |
| 182-190               | #region helix (right hand alpha)\                             |  |
| 14-16,45-50,          | #region beta sheet\   |  |
| 134-140,99-101        |   |  |
| 28-31,36-39,          |   |  |
| 168-178,155-162,      |   |  |
| 60-67,118-123,        |   |  |
| 108-111               |   |  |
| SUMMARY               | #region beta sheet  |  |
|                       | #length 191 #molecular-weight 21515 #checksum 8219            |  |
| Query Match           | 100.0%; Score 63; DB 5; Length 191;                           |  |
| Best Local Similarity | 100.0%; Pred. No. 5.18e-03;                                   |  |

Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 58 GTRVRAAI 66  
 1 GTRVRAAI 9

RESULT 2

ENTRY 1TUPB #type complete  
 TITLE tumor suppressor p53, chain B - human  
 PDB\_TITLE tumor suppressor p53 complexed with DNA  
 ORGANISM #formal\_name Homo sapiens #common\_name man  
 #note expressed in Escherichia coli  
 REFERENCE A66776  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
 #submission submitted to the Brookhaven Protein Data Bank, July 1995  
 #cross-references PDB:1TUP  
 REFERENCE A43072  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
 #journal Science (1994) 265:346-355  
 #title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE A49450  
 #authors Pavletich, N.P.; Chambers, K.A.; Pabo, C.O.  
 #journal Genes Dev. (1993) 7:2556-2564  
 #title The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots.  
 REFERENCE TN031795  
 #authors Vogelstein, B.; Kinzler, K.W.  
 #journal Cell (1992) 70:523  
 #title p53 function and dysfunction.  
 COMMENT Resolution: 2.2 angstroms  
 COMMENT Determination: X-ray diffraction  
 COMMENT R-value: 0.202

KEYWORDS antigen p53; complex; DNA; tumor suppressor

FEATURE 72-75 #region helix (right hand 3-10)\  
 82-86 #region helix (right hand alpha)\  
 183-192 #region helix (right hand alpha)\  
 15-17,46-51, #region beta sheet\  
 135-141,100-102  
 129-132,37-40,  
 169-179,156-163,  
 61-68,119-124,  
 109-112

SUMMARY #region beta sheet  
 #length 194 #molecular-weight 21830 #checksum 2852

Query Match 100.0%; Score 63; DB 5; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 5.18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 59 GTRVRAAI 67  
 1 GTRVRAAI 9

RESULT 3

ENTRY 1TSRB #type complete  
 TITLE p53 tumor suppressor, chain B - human  
 PDB\_TITLE p53 core domain in complex with DNA  
 ORGANISM #formal\_name Homo sapiens #common\_name man  
 #note expressed in Escherichia coli  
 REFERENCE A66760  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.; Pavletich, N.  
 #submission submitted to the Brookhaven Protein Data Bank, July 1995  
 #cross-references PDB:1TSR  
 REFERENCE A43072  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
 #journal Science (1994) 265:346-355  
 #title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.  
 COMMENT Resolution: 2.2 angstroms  
 COMMENT Determination: X-ray diffraction

KEYWORDS anti-oncogene; complex; DNA; DNA-binding protein; tumor suppressor

FEATURE 72-75 #region helix (right hand 3-10)\  
 82-86 #region helix (right hand alpha)\  
 183-192 #region helix (right hand alpha)\  
 15-17,46-51, #region beta sheet\  
 135-141,100-103  
 29-32,37-40,  
 169-179,156-163,  
 61-68,119-124,  
 109-112

SUMMARY #region beta sheet  
 #length 194 #molecular-weight 21830 #checksum 2852

Query Match 100.0%; Score 63; DB 5; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 5.18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 59 GTRVRAAI 67  
 1 GTRVRAAI 9

RESULT 4

ENTRY 1TUPC #type complete  
 TITLE tumor suppressor p53, chain C - human  
 PDB\_TITLE tumor suppressor p53 complexed with DNA  
 ORGANISM #formal\_name Homo sapiens #common\_name man  
 #note expressed in Escherichia coli  
 REFERENCE A66776  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
 #submission submitted to the Brookhaven Protein Data Bank, July 1995  
 #cross-references PDB:1TUP  
 REFERENCE A43072  
 #authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
 #journal Science (1994) 265:346-355  
 #title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE A49450  
 #authors Pavletich, N.P.; Chambers, K.A.; Pabo, C.O.  
 #journal Genes Dev. (1993) 7:2556-2564  
 #title The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots.  
 REFERENCE TN031798  
 #authors Vogelstein, B.; Kinzler, K.W.  
 #journal Cell (1992) 70:523  
 #title p53 function and dysfunction.  
 COMMENT Resolution: 2.2 angstroms  
 COMMENT Determination: X-ray diffraction  
 COMMENT R-value: 0.202

KEYWORDS antigen p53; complex; DNA; tumor suppressor

FEATURE 11-13 #region helix (right hand 3-10)\  
 72-75 #region helix (right hand alpha)\  
 83-87 #region helix (right hand alpha)\  
 184-191 #region helix (right hand alpha)\  
 16-19,47-52, #region beta sheet\  
 136-142,101-103  
 31-33,38-42,  
 170-181,157-164,  
 62-69,120-125,  
 110-113

SUMMARY #region beta sheet  
 #length 195 #molecular-weight 21917 #checksum 4657

Query Match 100.0%; Score 63; DB 5; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 5.18e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 60 GTRVRAAI 68  
 1 GTRVRAAI 9

RESULT 5  
ENTRY 1TSRC #type complete  
TITLE: p53 tumor suppressor, chain C - human  
PDB TITLE: p53 core domain in complex with DNA  
ORGANISM #formal\_name Homo sapiens #common\_name man  
#note expressed in Escherichia coli  
REFERENCE A66760  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.; Pavletich, N.  
#submission submitted to the Brookhaven Protein Data Bank, July 1995  
#cross-references PDB:1TSR  
REFERENCE A43072  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex:  
understanding tumorigenic mutations.  
COMMENT Resolution: 2.2 angstroms  
Determination: X-ray diffraction  
KEYWORDS anti-oncogene; complex; DNA; DNA-binding protein; tumor  
suppressor  
FEATURE  
11-13 #region helix (right hand 3-10)\\  
72-75 #region helix (right hand alpha)\\  
83-87 #region helix (right hand alpha)\\  
184-191 #region helix (right hand alpha)\\  
16-19,47-52, #region beta sheet\\  
136-142,101-103  
31-33,38-42, #region beta sheet\\  
170-181,157-164,  
62-69,120-125,  
110-113  
SUMMARY #length 195 #molecular-weight 21917 #checksum 4657  
#region beta sheet  
Query Match 100.0%; Score 63; DB 5; Length 195;  
Best Local Similarity 100.0%; Pred. No. 5.18e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 60 GTRVRAAI 68  
QY 1 GTRVRAAI 9  
RESULT 6  
ENTRY 1TRPA #type complete  
TITLE: tumor suppressor p53, chain A - human  
PDB TITLE: tumor suppressor p53 complexed with DNA  
ORGANISM #formal\_name Homo sapiens #common\_name man  
#note expressed in Escherichia coli  
REFERENCE A66776  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#submission submitted to the Brookhaven Protein Data Bank, July 1995  
#cross-references PDB:1TRP  
REFERENCE A43072  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex:  
understanding tumorigenic mutations.  
REFERENCE A49450  
#authors Pavletich, N.P.; Chambers, K.A.; Pabo, C.O.  
#journal Genes Dev. (1993) 7:2556-2564  
#title The DNA-binding domain of p53 contains the four conserved  
regions and the major mutation hot spots.  
REFERENCE TN031792  
#authors Vogelstein, B.; Kinzler, K.W.  
#journal Cell (1992) 70:523  
#title P53 function and dysfunction.  
COMMENT Resolution: 2.2 angstroms  
Determination: X-ray diffraction  
R-value: 0.202  
KEYWORDS antigen p53; complex; DNA; tumor suppressor  
FEATURE  
73-75 #region helix (right hand 3-10)\\  
84-87 #region helix (right hand alpha)\\  
185-194 #region helix (right hand alpha)\\  
17-19,48-53, #region beta sheet\\  
137-143,102-105  
31-34,39-42, #region beta sheet\\  
171-181,158-165,  
63-70,121-126,  
111-114  
SUMMARY #length 196 #molecular-weight 22004 #checksum 7058  
#region beta sheet  
Query Match 100.0%; Score 63; DB 5; Length 196;  
Best Local Similarity 100.0%; Pred. No. 5.18e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 61 GTRVRAAI 69  
QY 1 GTRVRAAI 9  
RESULT 8  
ENTRY S51648 #type complete  
TITLE: cellular tumor antigen p53 - bovine  
ALTERNATE\_NAMES tumor-suppressor protein p53  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change  
08-Sep-1997  
ACCESSIONS S51648  
REFERENCE S51648  
#authors Degludt, F.; Willems, L.; Burny, A.; Kettmann, R.  
#submission Submitted to the EMBL Data Library, September 1994  
#description Nucleotide sequence of the ovine p53 tumor-suppressor gene  
cDNA and its genomic organisation.  
#accession S51648  
#status preliminary  
#molecule\_type mRNA

17-19,48-53, #region beta sheet\\  
137-143,102-105  
31-34,39-42,  
171-181,158-165,  
63-70,121-126,  
111-114  
SUMMARY #length 196 #molecular-weight 22004 #checksum 7058  
#region beta sheet  
Query Match 100.0%; Score 63; DB 5; Length 196;  
Best Local Similarity 100.0%; Pred. No. 5.18e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 61 GTRVRAAI 69  
QY 1 GTRVRAAI 9  
RESULT 7  
ENTRY 1TSRA #type complete  
TITLE: p53 tumor suppressor, chain A - human  
PDB TITLE: p53 core domain in complex with DNA  
ORGANISM #formal\_name Homo sapiens #common\_name man  
#note expressed in Escherichia coli  
REFERENCE A66760  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.; Pavletich, N.  
#submission submitted to the Brookhaven Protein Data Bank, July 1995  
#cross-references PDB:1TSR  
REFERENCE A43072  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex:  
understanding tumorigenic mutations.  
COMMENT Resolution: 2.2 angstroms  
Determination: X-ray diffraction  
KEYWORDS anti-oncogene; complex; DNA; DNA-binding protein; tumor  
suppressor  
FEATURE  
73-75 #region helix (right hand 3-10)\\  
84-87 #region helix (right hand alpha)\\  
185-194 #region helix (right hand alpha)\\  
17-19,48-53, #region beta sheet\\  
137-143,102-105  
31-34,39-42, #region beta sheet\\  
171-181,158-165,  
63-70,121-126,  
111-114  
SUMMARY #length 196 #molecular-weight 22004 #checksum 7058  
#region beta sheet  
Query Match 100.0%; Score 63; DB 5; Length 196;  
Best Local Similarity 100.0%; Pred. No. 5.18e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 61 GTRVRAAI 69  
QY 1 GTRVRAAI 9  
RESULT 8  
ENTRY S51648 #type complete  
TITLE: cellular tumor antigen p53 - bovine  
ALTERNATE\_NAMES tumor-suppressor protein p53  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change  
08-Sep-1997  
ACCESSIONS S51648  
REFERENCE S51648  
#authors Degludt, F.; Willems, L.; Burny, A.; Kettmann, R.  
#submission Submitted to the EMBL Data Library, September 1994  
#description Nucleotide sequence of the ovine p53 tumor-suppressor gene  
cDNA and its genomic organisation.  
#accession S51648  
#status preliminary  
#molecule\_type mRNA

##residues 1-386 ##label DEQ  
##cross-references EMBL:X81704; NID:g602332; PID:g602333  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;  
phosphoprotein; transcription regulation; tumor suppressor;  
zinc

FEATURE 1  
168,171,231,235 #binding\_site zinc (Cys, His, Cys, Cys) #status  
#predicted  
385 #binding\_site phosphoryl-RNA (Ser) (covalent) #status  
#predicted

SUMMARY 1  
#length 386 #molecular-weight 43355 #checksum 7025

Query Match 100.0%; Score 63; DB 2; Length 386;  
Best Local Similarity 100.0%; Pred. No. 5.18e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 146 GTRVRAAI 154  
|||  
OY 1 GTRVRAAI 9

RESULT 9  
ENTRY JC6193 #type complete  
TITLE tumor suppressor p53 - rabbit  
ORGANISM #formal\_name Oryctolagus cuniculus #common\_name domestic  
rabbit  
DATE 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change  
08-Sep-1997

ACCESSIONS JC6193  
REFERENCE JC6193  
#authors Le Goas, F.; May, P.; Ronco, P.; de Fromental, C.C.  
#journal Gene (1997) 185:169-173  
#title cDNA cloning and immunological characterization of rabbit  
p53

GENETICS  
#accession JC6193  
#molecule\_type mRNA  
#residues 1-391 ##label LEE  
##cross-references EMBL:X90592; NID:g1532043; PID:e194962; PID:g1532044

CLASSIFICATION p53  
KEYWORDS #superfamily cellular tumor antigen p53  
SUMMARY #length 391 #molecular-weight 43435 #checksum 4367

Query Match 100.0%; Score 63; DB 2; Length 391;  
Best Local Similarity 100.0%; Pred. No. 5.18e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 151 GTRVRAAI 159  
|||  
OY 1 GTRVRAAI 9

RESULT 10  
ENTRY S02192 #type complete  
TITLE cellular tumor antigen p53 - rat  
ALTERNATE\_NAMES gene p53 protein; nuclear oncoprotein p53  
ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
DATE 18-Oct-1988 #sequence\_revision 18-Oct-1989 #text\_change  
08-Sep-1997  
S02192; S41149  
S02192  
#authors Soussi, T.; de Fromental, C.C.; Breugnot, C.; May, E.  
#journal Nucleic Acids Res. (1988) 16:11384  
#title Nucleotide sequence of a cDNA encoding the rat p53 nuclear  
oncoprotein.  
#cross-references MUID:89083585  
#accession S02192  
#molecule\_type mRNA  
#residues 1-391 ##label SOU  
##cross-references EMBL:X13058; NID:g56828; PID:g56829  
S41149

##authors Hulla, J.E.; Schneider, R.P.  
#journal Nucleic Acids Res. (1993) 21:713-717  
#title Structure of the rat p53 tumor suppressor gene.  
#accession S41149  
#status preliminary; nucleic acid sequence not shown;  
translation not shown

##molecule\_type DNA  
##residues 1-173, 'N', 175-391 ##label HTL  
##cross-references EMBL:L07909  
##note the nucleotide sequence was submitted to the EMBL Data  
Library, December 1992

GENETICS  
#introns 25/2; 32/3; 123/3; 185/1; 259/2; 305/1; 329/3; 365/2  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;  
nucleus; phosphoprotein; transcription regulation; tumor  
suppressor; zinc

FEATURE 1  
174,177,236,240 #binding\_site zinc (Cys, His, Cys, Cys) #status  
#predicted  
390 #binding\_site phosphoryl-RNA (Ser) (covalent) #status  
#predicted

SUMMARY 1  
#length 391 #molecular-weight 43451 #checksum 7105

Query Match 100.0%; Score 63; DB 2; Length 391;  
Best Local Similarity 100.0%; Pred. No. 5.18e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 152 GTRVRAAI 160  
|||  
OY 1 GTRVRAAI 9

RESULT 11  
ENTRY JC6176 #type complete  
TITLE tumor suppressor protein p53 - Chinese hamster  
ORGANISM #formal\_name Citicellus griseus #common\_name Chinese hamster  
DATE 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change  
08-Sep-1997

ACCESSIONS JC6176  
REFERENCE JC6176  
#authors Lee, H.; Larner, J.M.; Hamlin, J.L.  
#journal Gene (1997) 184:177-183  
#title Cloning and characterization of Chinese hamster p53 cDNA.  
#contents liver  
#accession JC6176  
#molecule\_type mRNA  
#residues 1-393 ##label LEE  
##cross-references GB:U50395; NID:g1842229; PID:g1842230  
COMMENT This protein is a multimer, it plays the central role in a complex  
DNA damage-sensing network, it binds to replication factor and  
TATA-binding protein, and affects DNA replication, transcription,  
and recombination by protein/protein interactions.

GENETICS  
#gene p53  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS #liver: tumor  
SUMMARY #length 393 #molecular-weight 43362 #checksum 4043

Query Match 100.0%; Score 63; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 5.18e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAAI 162  
|||  
OY 1 GTRVRAAI 9

RESULT 12  
ENTRY DNHU53 #type complete  
TITLE cellular tumor antigen p53 - human  
ALTERNATE\_NAMES cellular phosphoprotein p53; oncoprotein p53; transformation  
suppressor p53; tumor suppressor p53

```

ORGANISM      #formal_name Homo sapiens #common_name man
DATE          05-Oct-1988 #sequence_revision 18-Nov-1994 #text_change
ACCESSIONS    A25224; A43073; JT0436; S40773; S42669; A22837; A55060;
              A25397; B25397; S42452; S42453; I38082; I38083; I38084;
              I38085; I38086; I38087; I38088; I38089; I38090; I38091;
              I38092; I38093; A44905; I58354; I78850; S60153
              A25224
REFERENCE      #authors      Lamb, P.; Crawford, L.
              #journal      Mol. Cell. Biol. (1986) 6:1379-1385
              #title        Characterization of the human p53 gene.
              #cross-references MUID:87064416
              #accession     A25224
              #molecule_type DNA
              ##residues     1-393 ##label LAM
              ##cross-references EMBL:X01405; GB:M13121; GB:N00032; NID:q189460;
              PDB:g386994
REFERENCE      JT0436
              #authors      Buchan, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.;
              Georgiev, G.P.
              #journal      Gene (1988) 70:245-252
              #title        A variation in the structure of the protein-coding region of
              the human p53 gene.
              #cross-references MUID:89108008
              #accession     A43073
              ##molecule_type DNA
              ##residues     1-393 ##label BUC
              #note          this 72-Avg allele appears to be about 5 times more
              frequent than the 72-Pro allele
              #accession     JT0436
              ##molecule_type DNA
              ##residues     1-71, 'P', 73-393 ##label BU2
              ##cross-references EMBL:M2889; NID:q189474; PID:q189476
              #note          this 72-Pro allele was found in both normal and
              malignant cell lines
REFERENCE      S40773
              #authors      Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.
              #journal      submitted to the EMBL Data Library, August 1990
              #accession     S40773
              ##molecule_type DNA
              ##residues     1-393 ##label CHU
              ##cross-references EMBL:X54156; NID:g35213; PID:g35214
              #accession     S42669
              #authors      Matlashewski, G.; Lamb, P.; Pim, D.; Peacock, J.; Crawford,
              L.; Benchimol, S.
              #journal      EMBO J. (1984) 3:3257-3262
              #title        Isolation and characterization of a human p53 cDNA clone:
              expression of the human p53 gene.
              #accession     S42669
              ##molecule_type mRNA
              ##residues     101-393 ##label MKI
              ##cross-references EMBL:X01405; NID:g35215; PID:g642241
REFERENCE      A22837
              #authors      Zakut-Houri, R.; Bienz-Tadmor, B.; Glyvol, D.; Oren, M.
              #journal      EMBO J. (1985) 4:1251-1255
              #title        Human p53 cellular tumor antigen: cDNA sequence and
              expression in COS cells.
              #cross-references MUID:85230577
              #accession     A22837
              ##molecule_type mRNA
              ##residues     1-71, 'P', 73-393 ##label ZAK
              ##cross-references EMBL:X02469; EMBL:M60950; NID:g35209; PID:g35210
REFERENCE      A55060
              #authors      Harlow, E.; Williamson, N.M.; Ralston, R.; Helfman, D.M.;
              Adams, T.E.
              #journal      Mol. Cell. Biol. (1985) 5:1601-1610
              #title        Molecular cloning and in vitro expression of a cDNA clone for
              human cellular tumor antigen p53.
              #accession     A55060
              ##molecule_type mRNA
              ##residues     1-71, 'P', 73-272, 'H', 274-393 ##label HA3
              ##cross-references GB:K03199; NID:q189478; PID:q189479
              #experimental_source clone PR4-2, cell line A431

```

---

```

REFERENCE      A93086
              #authors      Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.;
              Arari, N.; Rotter, V.
              #journal      Mol. Cell. Biol. (1986) 6:4650-4656
              #title        Molecular basis for heterogeneity of the human p53 protein.
              #cross-references MUID:87089826
              #accession     A25397
              ##molecule_type mRNA
              ##residues     1-78, 'T', 80-393 ##label HAR
              ##cross-references EMBL:M44694; NID:g339813; PID:g339814
              #experimental_source clone p53-H-1, transformed hybridoma SV-80 cell
              line
              #accession     B25397
              ##molecule_type mRNA
              ##residues     1-71, 'P', 73-78, 'T', 80-393 ##label HA2
              ##cross-references EMBL:M44695; NID:g339815; PID:g339816
              #experimental_source clone p53-H-19, transformed hybridoma SV-80 cell
              line
REFERENCE      S42452
              #authors      Matlashewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider,
              J.; Crawford, L.V.
              #journal      Mol. Cell. Biol. (1987) 7:961-963
              #title        Primary structure polymorphism at amino acid residue 72 of
              human p53.
              #accession     S42452
              ##molecule_type mRNA; DNA
              ##residues     66-71, 'P', 73-79 ##label MK2
              #note          #experimental_source clone lambda C113
              72-Cys was also found, and appears to represent a
              polymorphism
              #accession     S42453
              ##molecule_type mRNA; DNA
              ##residues     66-79 ##label MAT
              #experimental_source clone J6K
REFERENCE      I38082
              #authors      Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.;
              Crook, T.
              #journal      EMBO J. (1991) 10:2879-2887
              #title        p53 is frequently mutated in Burkitt's lymphoma cell lines.
              #cross-references MUID:92007731
              #accession     I38082
              ##status        translated from GB/EMBL/DBJ
              ##molecule_type mRNA
              ##residues     1-189, 'L', 'S', 'E', 'K', 'E', 'I', 'C', 'V', 'S', 'I', 'M', 'T', 'E', 'T', 'F', 'D', 'I', 'V', 'M', 'C', 'P', 'S', 'R', 'L', 'A', 'L', 'T',
              'V', 'P', 'S', 'T', 'T', 'C', 'T', 'V', 'P', 'A', 'M', 'A', ' ' ##label F01
              #cross-references EMBL:X60010; NID:g506432; PID:g506433
              #note          deletion of a C nucleotide causes a frameshift at
              position 566
              #accession     I38083
              ##status        translated from GB/EMBL/DBJ
              ##molecule_type mRNA
              ##residues     1-192, 'R', 194-393 ##label F02
              #cross-references EMBL:X60011; NID:g506434; PID:g506435
              #accession     I38084
              ##status        translated from GB/EMBL/DBJ
              ##molecule_type mRNA
              ##residues     1-393 ##label F03
              #cross-references EMBL:X60012; NID:g506436; PID:g506437
              #accession     I38085
              ##status        translated from GB/EMBL/DBJ
              ##molecule_type mRNA
              ##residues     1-245, 'T', 247-393 ##label F04
              #cross-references EMBL:X60013; NID:g506438; PID:g506439
              #accession     I38086
              ##status        translated from GB/EMBL/DBJ
              ##molecule_type mRNA
              ##residues     1-236, 'T', 238-393 ##label F05
              #cross-references EMBL:X60014; NID:g506440; PID:g506441
              #accession     I38087
              ##status        translated from GB/EMBL/DBJ
              ##molecule_type mRNA
              ##residues     1-247, 'Q', 249-393 ##label F06
              #cross-references EMBL:X60015; NID:g506442; PID:g506443

```

```

#accession I38088
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-71, 'P', 73-237, 'Y', 238-393 ##label F07
#cross-references EMBL:X60016; NID:g506444; PID:g506445
#accession I38089
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-247, 'O', 249-393 ##label F08
#cross-references EMBL:X60017; NID:g506446; PID:g506447
#accession I38090
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-71, 'P', 73-162, 'H', 164-393 ##label F09
#cross-references EMBL:X60018; NID:g506448; PID:g506449
#accession I38091
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-212, 'O', 214-393 ##label F10
#cross-references EMBL:X60019; NID:g506450; PID:g506451
#accession I38092
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-253, 'D', 255-393 ##label F11
#cross-references EMBL:X60020; NID:g506452; PID:g506453
#note all sequences submitted to the EMBL/GenBank/DBJ
databases June 1991

REFERENCE
#authors Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.
#journal Nucleic Acids Res. (1991) 19:6977
#title An Alu polymorphism intragenic to the TP53 gene.
#cross-references MIMD:92107726
#accession I38093
#status translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-393 ##label RE2
#cross-references EMBL:X54156; NID:g35213; PID:g35214
#accession A44905
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;
Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.;
#journal Cancer Res. (1991) 51:5800-5805
#title p53 gene mutations in gastric cancer metastases and in
gastric cancer cell lines derived from metastases.
#cross-references MIMD:92034678
#accession A44905
#molecule_type DNA
#residues 246-247, 'W', 249-250 ##label YAM
#cross-references GB:S63157; NID:g237829; PID:g237830
#note sequence extracted from NCBI backbone (NCBIN:63157,
sequence remainder of annotations omitted.

Note: remainder of annotations omitted.

Query Match 100.0%; Score 63; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 5.18e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAAI 162
OY 1 GTRVRAAI 9

```

```

#title The cDNA cloning and immunological characterization of
hamster p53
#cross-references MIMD:92210007
#accession JH0633
#molecule_type mRNA
#residues 1-396 ##label LEG
#cross-references GB:M75144; NID:g191414; PID:g191415
#experimental_source kidney, strain MPI

GENETICS
#gene p53
CLASSIFICATION
#superfamily cellular tumor antigen p53
#apoptosis; cell division control; DNA binding; homotrimer;
#nucleus; phosphoprotein; transcription regulation; tumor
#suppressor; zinc

FEATURE
179,182,241,245 #binding_site zinc (Cys, His, Cys, Cys) #status
395 #predicted\
#binding_site phosphoryl-RNA (Ser) (covalent) #status
395 #predicted

SUMMARY
#length 396 #molecular_weight 43631 #checksum 6617

Query Match 100.0%; Score 63; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 5.18e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 157 GTRVRAAI 165
OY 1 GTRVRAAI 9

RESULT 14
ENTRY S38824 #type complete
#cellular tumor antigen p53, alternative splice form - mouse
#formal_name Mus musculus #common_name house mouse
#organism Mus musculus
#date 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
25-Oct-1996
#accessions S38824; S35478
#cross-references S38822
#authors Arsl, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:3232-3239
#title Immunologically distinct p53 molecules generated by
alternative splicing.
#accession S38824
#molecule_type mRNA
#residues 1-381 ##label ARA
#cross-references S35478
#authors Han, K.A.; Kulesz-Martin, M.F.
#journal Nucleic Acids Res. (1992) 20:1979-1981
#title Alternatively spliced p53 RNA in transformed and normal cells
of different tissue types.
#accession S35478
#status nucleic acid sequence not shown; translation not shown
#molecule_type RNA
#residues 1-381 ##label HAN
#cross-references EMBL:M13874; NID:g200202; PID:g200203
#note the nucleotide sequence was submitted to the EMBL Data
Library, July 1988

COMMENT This sequence, produced by alternative splicing of the tenth
intron, lacks the carboxyl-terminal sequence necessary for
covalent attachment of RNA. The function of this minor splice
form is not known.

CLASSIFICATION #superfamily cellular tumor antigen p53
#alternative splicing; phosphoprotein
KEYWORDS
1-44 #domain transcription activation #status predicted
#label TRA\
16-26 #region conserved region I\
99-289 #domain DNA-binding core #status predicted #label DBC\
108-121 #region LI loop\
114-139 #region conserved region II\
160-192 #region I2 loop\
168-178 #region conserved region III\

```



```

231-252      #region conserved region IV\
233-248      #region L3 loop\
267-283      #region conserved region V\
313-319      #region nuclear location signal\
319-357      #region tetramer association\
7,9,12,18,23,37 #binding-site phosphate (Ser) (covalent) #status
173,176,235,239 #binding-site zinc (Cys, His, Cys, Cys) #status
312          #predicted\
              #binding-site phosphate (Ser) (covalent) (by cdc2
              #kinase) #status predicted
SUMMARY      #length 381 #molecular-weight 42498 #checksum 8703
Query Match  95.2% Score 60; DB 2; Length 381;
Best Local Similarity 88.9%; Pred. No. 2,55e-02;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 151 GSRVRAAI 159
QY 1 GTRVRAAI 9

RESULT 15
ENTRY      DNMS53      #type complete
TITLE      cellular tumor antigen p53 - mouse
ALTERNATE_NAMES
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        28-Aug-1985 #sequence_revision 04-Oct-1996 #text_change
05-Sep-1997
ACCESSIONS  A22739; S06336; A02684; S38822; S38823; I48703
REFERENCE   A22739
#authors    Bienz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.
#journal    EMBO J. (1984) 3:2179-2183
#cross-references MIMD:85027173
#accession  A22739
#molecule-type DNA
##residues 1-134,'V',136-390 ##label BIE
REFERENCE   S06336
#authors    Chumakov, P.M.
#journal    Bioorg. Khim. (1987) 13:1691-1694
#title      Primary structure of DNA complementary to murine oncoprotein
p53 mRNA.
#cross-references MIMD:88221682
#accession  S06336
#status     not compared with conceptual translation
#molecule-type mRNA
##residues 1-134,'V',136-390 ##label CHU
REFERENCE   A02684
#authors    Zakut-Houri, R.; Oren, M.; Bienz, B.; Lavie, V.; Hazum, S.;
Givol, D.
#journal    Nature (1983) 306:594-597
#title      A single gene and a pseudogene for the cellular tumour
antigen p53.
#cross-references MIMD:84068204
#accession  A02684
#molecule-type mRNA
##residues 1-159,'H',161-167,'G',169-233,'T',235-390 ##label ZAK
REFERENCE   S38822
#authors    Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#journal    Mol. Cell. Biol. (1986) 6:3232-3239
#title      Immunologically distinct p53 molecules generated by
alternative splicing.
#accession  S38822
#status     preliminary
#molecule-type mRNA
##residues 1-390 ##label ARA
#cross-references EMBL:M13872; NID:9200198; PTD:9200199
#accession  S38823
#status     preliminary
#molecule-type mRNA
##residues 1-167,'G',169-233,'T',235-390 ##label AR2
#cross-references EMBL:M13873

```

```

REFERENCE   I48703
#authors    Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.
#journal    Nucleic Acids Res. (1984) 12:5609-5626
#title      Cloning and expression analysis of full length mouse CDNA
sequences encoding the transformation associated protein
p53.
#cross-references MIMD:84272240
#accession  I48703
#status     preliminary; translated from GB/EMBL/DBJ
#molecule-type mRNA
##residues 1-47,'R',49-78,'OW',82-390 ##label RES
##residues 1-47,'R',49-78,'OW',82-390 ##label RES
#cross-references EMBL:X00741; NID:953570; PTD:953571
COMMENT      This DNA-binding protein plays an essential role in the regulation
of cell division, as it is required for the transition from phase
G0 to G1 of the cell cycle.
COMMENT      The tetramer association region may exhibit a beta-turn,
beta-sheet, beta-turn, alpha-helix motif.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS      apoptosis; cell division control; DNA binding; homotetramer;
phosphoprotein; transcription regulation; tumor suppressor;
zinc
FEATURE      1-44      #domain transcription activation #status predicted
16-26      #label TRA\
99-289      #region conserved region I\
108-121      #region DNA-binding core #status predicted #label DBC\
114-139      #region L1 loop\
160-192      #region conserved region II\
168-178      #region L2 loop\
231-252      #region conserved region III\
233-248      #region conserved region IV\
267-283      #region L3 loop\
313-319      #region conserved region V\
319-357      #region nuclear location signal\
7,9,12,18,23,37 #binding-site phosphate (Ser) (covalent) #status
173,176,235,239 #binding-site zinc (Cys, His, Cys, Cys) #status
312          #predicted\
              #binding-site phosphate (Ser) (covalent) (by cdc2
              #kinase) #status predicted\
              #binding-site phosphoryl-RNA (Ser) (covalent) #status
389          #predicted
SUMMARY      #length 390 #molecular-weight 43458 #checksum 1260
Query Match  95.2% Score 60; DB 1; Length 390;
Best Local Similarity 88.9%; Pred. No. 2,55e-02;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 151 GSRVRAAI 159
QY 1 GTRVRAAI 9

Search completed: Fri Sep 11 13:36:57 1998
Job time : 18 secs.

```







FT DOMAIN 300 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 381 381 PHOSPHORYLATION (BY SIMILARITY).  
 SO SEQUENCE 382 AA; 42809 MW; OCB99A00 CRC32;  
 Query Match 100.0%; Score 63; DB 1; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 2,60e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 142 GTRVRAAI 150  
 OY 1 GTRVRAAI 9

RESULT 5 STANDARD; PRT; 386 AA.  
 ID P53 BOVIN  
 AC Q29628;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53  
 OS BOS TAURUS (BOVINE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; ARTIODACTYLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RX MEDLINE; 95352829.  
 RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
 RL DNA SEQ. 5:261-264(1995).  
 RN [2]  
 RP SEQUENCE OF 13-386 FROM N.A.  
 RC STRAIN=HOLSTEIN; TISSUE=THYMUS;  
 RX MEDLINE; 96401400.  
 RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;  
 RL VET. IMMUNOL. IMMUNOPATHOL. 52:53-63(1996).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL; X81704; G602333; -;  
 DR EMBL; D49825; G1729419; -;  
 DR PROSITE; PS00348; P53; 1.  
 KM ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 39  
 FT MOD\_RES 304 316 ASP/GLU-RICH (ACIDIC).  
 FT MOD\_RES 385 385 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT CONFLICT 380 380 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 380 380 R -> T (IN REF. 2).  
 SO SEQUENCE 386 AA; 43255 MW; 0322BF3D CRC32;  
 Query Match 100.0%; Score 63; DB 1; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 2,60e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 146 GTRVRAAI 154  
 OY 1 GTRVRAAI 9

RESULT 6 STANDARD; PRT; 391 AA.  
 ID P53 RAT  
 AC Q29628;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR P53.  
 OS RATTOUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE; 89083585.  
 RA SOUSSET T.;  
 RL NUCLEIC ACIDS RES. 16:11384-11384(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE; 93181268.  
 RA HULLA J.E., SCHNEIDER R.P.;  
 RL NUCLEIC ACIDS RES. 21:713-717(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY;  
 RA MATRUPALA S.P.;  
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL; X13058; G56829; -;  
 DR EMBL; L07910; G205952; -;  
 DR EMBL; L07904; G205952; JOINED.  
 DR EMBL; L07905; G205952; JOINED.  
 DR EMBL; L07906; G205952; JOINED.  
 DR EMBL; L07907; G205952; JOINED.  
 DR EMBL; L07908; G205952; JOINED.  
 DR EMBL; L07909; G205952; JOINED.  
 DR EMBL; U90328; G1938365; -;  
 DR PIR; S02192; S02182.  
 DR HSSP; P04637; IPES.  
 DR PROSITE; PS00348; P53; 1.  
 KM ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 76  
 FT MOD\_RES 277 391 ASP/GLU-RICH (ACIDIC).  
 FT MOD\_RES 309 321 HIGHLY BASIC AND MAY BE INVOLVED IN  
 FT MOD\_RES 330 390 INTERACTION WITH DNA.  
 FT VARIANT 103 103 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT VARIANT 256 256 PHOSPHORYLATION (BY SIMILARITY).  
 FT VARIANT 174 174 G -> S.  
 FT CONFLICT 174 174 E -> G.  
 FT CONFLICT 174 174 C -> W (IN REF. 2).  
 SO SEQUENCE 391 AA; 43451 MW; E0114C18 CRC32;  
 Query Match 100.0%; Score 63; DB 1; Length 391;  
 Best Local Similarity 100.0%; Pred. No. 2,60e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 152 GTRVRAAI 160  
 OY 1 GTRVRAAI 9

```

RESULT 7
ID P53_RABIT STANDARD: PRT: 391 AA.
AC 095330: 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; LAGOMORPHA.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 97208869.
RX LE GOAS F., MAY P., RONCO P., CARON DE FROMENTEL C.;
  GENE 185:169-173(1997).
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
  GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
  CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN
  TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
  TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
  BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
  THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
  APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
  BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
  EXPRESSION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
  OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
  IN MANY TYPES OF CANCER.
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.
DR EMBL: X90592: E194962: -.
DR PROSITE: PS00348: P53: 1.
DR ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
  NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.
KM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.
FT DOMAIN 1 70
  NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD.RES 308 321
  PHOSPHORYLATION (BY SIMILARITY).
FT MOD.RES 391 AA: 43435 MW; 30A36172 CRC32:
  SEQUENCE
  Query Match 100.0%; Score 63; DB 1; Length 391;
  Best Local Similarity 100.0%; Pred. No. 2.60e-04;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 151 GTRVAMAI 159
  1111111111
  QY 1 GTRVAMAI 9

RESULT 8
ID P53_CRIGR STANDARD: PRT: 393 AA.
AC 009185: 064397; P97258; P97788;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53 OR P53.
OS CRICETULUS GRISEUS (CHINESE HAMSTER).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RA CHUNG W., MI L.J., BOORSTEIN R.J.;
  SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-LAYER:
  MEDLINE: 97183659.
RX LEE H., LARNER J.M., HAMLIN J.L.;
  GENE 184:177-183(1997).
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
  GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL

```

```

CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
  OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
  IN MANY TYPES OF CANCER.
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.
DR EMBL: Y08900: E303876: -.
DR EMBL: Y08901: E303863: -.
DR EMBL: U50395: G1842230: -.
DR PROSITE: PS00348: P53: 1.
KM ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
  NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.
FT DOMAIN 1 74
  ASP/GLU-RICH (ACIDIC).
FT DOMAIN 75 150
  HYDROPHOBIC.
FT DOMAIN 316 390
  HIGHLY BASIC AND MAY BE INVOLVED IN
  INTERACTION WITH DNA.
FT DOMAIN 311 323
  NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD.RES 392 392
  PHOSPHORYLATION (BY SIMILARITY).
FT VARIANT 133 133
  L -> Q (IN CELL LINE V79-4).
FT VARIANT 135 135
  C -> W (IN CELL LINE V79-4).
FT CONFLICT 103 103
  Y -> F (IN REF. 2).
FT SEQUENCE 393 AA: 43378 MW; 402EB149 CRC32:
  Query Match 100.0%; Score 63; DB 1; Length 393;
  Best Local Similarity 100.0%; Pred. No. 2.60e-04;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVAMAI 162
  1111111111
  QY 1 GTRVAMAI 9

RESULT 9
ID P53_HUMAN STANDARD: PRT: 393 AA.
AC P04637:
DT 13-AUG-1987 (REL. 05, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).
GN TP53.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85230577.
RA ZAKUT-HOURI R., BIENZ-TADMOR B., GIVOL D., OREN M.;
  EMO J. 4:1251-1255(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87064416.
RA LAMB P., CRAWFORD L.;
  MOL. CELL. BIOL. 6:1379-1385(1986).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85267676.
RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;
  MOL. CELL. BIOL. 5:1601-1610(1985).
RN [4]
RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.
RX MEDLINE: 87089826.
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,
  ROTTNER V.;
  MOL. CELL. BIOL. 6:4650-4656(1986).
RN [5]
RP SEQUENCE FROM N.A.

```

RX MEDLINE: 89108008.  
 RA BUCHANAN V.L., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P.,  
 RA GEORGIEV G.P.;  
 RL GENE 70:245-252(1988).  
 RN [6]  
 RP SEQUENCE OF 101-393 FROM N.A.  
 RX MEDLINE: 85126934.  
 RA MATLASHENSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
 RA BENCHIMOL S.;  
 RL EMBO J. 3:3257-3262(1984).  
 RN [7]  
 RP NOCLEAR LOCALIZATION SIGNAL.  
 RX MEDLINE: 90191730.  
 RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
 RL ONCOGENE 5:423-426(1990).  
 RN [8]  
 RP STRUCTURE BY NMR OF 319-360.  
 RX MEDLINE: 94294808.  
 RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
 RA APPELLA E., GRONENBORN A.M.;  
 RL SCIENCE 265:386-391(1994).  
 RN [9]  
 RP STRUCTURE BY NMR OF 325-355.  
 RX MEDLINE: 95292092.  
 RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
 RL NAT. STRUCT. BIOL. 1:877-890(1994).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
 RX MEDLINE: 94294806.  
 RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
 RL SCIENCE 265:346-355(1994).  
 RN [11]  
 RP REVIEW.  
 RX MEDLINE: 94090335.  
 RA HARRIS C.C.;  
 RL SCIENCE 262:1980-1981(1993).  
 RN [12]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE: 91289156.  
 RA HOULSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
 RL SCIENCE 253:49-53(1991).  
 RN [13]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE: 96271983.  
 RA DE VRIES E.M.G., RICKE D.O., DE VRIES T.N., HARTMANN A., BLASZK H.,  
 RA LIAO D., SOUSSI T., KOVACH J.S., SOMMER S.S.;  
 RL HUM. MUTAT. 7:202-213(1996).  
 RN [14]  
 RP VARIANT ARG-72.  
 RX MEDLINE: 91153807.  
 RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
 RL HUM. GENET. 86:369-370(1991).  
 RN [15]  
 RP VARIANT LFS THR-133.  
 RX MEDLINE: 92034774.  
 RA LAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
 RL CANCER RES. 51:6385-6387(1991).  
 RN [16]  
 RP VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
 RX MEDLINE: 91057657.  
 RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
 RA KIM D.H., KASSEL J., GRAY M.A., BISCHOFF F.Z., TAINSKY M.A.,  
 RA FRIEND S.H.;  
 RL SCIENCE 250:1233-1238(1990).  
 RN [17]  
 RP VARIANT LFS ASP-245.  
 RX MEDLINE: 91080929.  
 RA SRIYASTAVA S., ZOU Z., PIROLOLO K., BLATTNER W., CHANG E.H.;  
 RL NATURE 348:747-749(1990).  
 RN [18]  
 RP VARIANT LFS LEU-272.  
 RX MEDLINE: 92147883.  
 RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,

RA POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
 RA KNUTSEN T., MINNA J.D.;  
 RL J. CLIN. INVEST. 89:640-647(1992).  
 RN [19]  
 RP VARIANTS LFS HIS-273 AND VAL-325.  
 RX MEDLINE: 92228023.  
 RA MALKIN D., JOLLY K.W., BARBER N., LOOK A.T., FRIEND S.H.,  
 RA GEBHARDT M.C., ANDERSEN T.I., BORESEN A.-L., LI F.P., GABER J.,  
 RA STRONG L.C.;  
 RL NEW ENGL. J. MED. 326:1309-1315(1992).  
 RN [20]  
 RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
 RX MEDLINE: 90295284.  
 RA BARTEK J., IGGO R., GANNON J., LANE D.P.;  
 RL ONCOGENE 5:893-899(1990).  
 RN [21]  
 RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
 RX MEDLINE: 91017544.  
 RA RODRIGUES N.R., ROMAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
 RA GANNON J.V., LANE D.P.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 87:7555-7559(1990).  
 RN [22]  
 RP VARIANTS COLORECTAL CANCER MUTATIONS.  
 RX MEDLINE: 91282784.  
 RA ISHIOKA C., SATO T., GANOH M., SUZUKI T., SHIBATA H., KANAMARU R.,  
 RA WAKI A., YAMAZAKI T.;  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 177:901-906(1991).  
 RN [23]  
 RP VARIANTS OESOPHAGUS TUMORS L-152; A-155; H-175; F-176 AND H-273.  
 RX MEDLINE: 91330175.  
 RA CASSON A.G., MUKHOPADHYAY T., CLEARY K.R., RO J.Y., LEVIN B.,  
 RA ROTH J.A.;  
 RL CANCER RES. 51:4495-4499(1991).  
 RN [24]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN CHINA.  
 RX MEDLINE: 91187113.  
 RA HSU I.C., METCALF R.A., SUN T., WELSH J.A., WANG N.J., HARRIS C.C.;  
 RL NATURE 350:427-428(1991).  
 RN [25]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN SOUTH AFRICA.  
 RX MEDLINE: 91187114.  
 RA BRESSAC B., KEM M., WANDS J., OZTURK M.;  
 RL NATURE 350:428-431(1991).  
 RN [26]  
 RP VARIANTS IN ANOGENITAL CARCINOMAS.  
 RX MEDLINE: 93010989.  
 RA CROOK T., VOUSDEN K.H.;  
 RL EMBO J. 11:3935-3940(1992).  
 RN [27]  
 RP VARIANT WITH DUPLICATION OF PRO-177-HIS-178.  
 RX MEDLINE: 93265016.  
 RA BHARIA K., GUTTEREZ M.I., MAGRATH I.T.;  
 RL HUM. MOL. GENET. 1:207-208(1992).  
 RN [28]  
 RP VARIANTS IN BURKITT'S LYMPHOMAS.  
 RX MEDLINE: 93064692.  
 RA DUTHU A., DEBUIRE B., ROMANO J.W., EHRLHART J.C., FISCELLA M., MAY E.,  
 RA APPELLA E., MAY P.;  
 RL ONCOGENE 7:2161-2167(1992).  
 RN [29]  
 RP VARIANT NASOPHARYNGEAL CARCINOMA THR-280.  
 RX MEDLINE: 92335329.  
 RA SUN Y., HEGAMAYER G., HENG Y.-Y., HILDESHEIM A., CHEN J.-Y., CAO Y.,  
 RA YAO K.-T., COLBURN N.H.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:6516-6520(1992).  
 RN [30]  
 RP VARIANTS IN COLON TUMORS.  
 RX MEDLINE: 93330562.  
 RA HAMELIN R., JEGO N., LAURENT-PUIG P., VIDAUD M., THOMAS G.;  
 RL ONCOGENE 8:2213-2220(1993).  
 RN [31]  
 RP CHARACTERIZATION OF VARIANT ALA-143.  
 RX MEDLINE: 94283378.

RA ZHANG W., GUO X.-Y., HU G.-Y., LIU W.-B., SHAY J.W., DEISSEROTH A.B.;  
 RN EMO J. 13:2535-2544(1994).  
 [32]  
 RA VARIANTS LFS HIS-175; ARG-193; GLN-248; CYS-273 AND TYR-275.  
 RX MEDLINE; 95193787.  
 RA FERDUNG T., BARBIER N., YAN Y.-X., GARDER J.E., DREYFUS M.,  
 RA FRAUMENI J.F. JR., LI F.P., FRIEND S.H.;  
 RN AM. J. HUM. GENET. 56:608-615(1995).  
 [33]  
 RA VARIANTS LFS HIS-175.  
 RX MEDLINE; 96423319.  
 RA VANLIER J.M., MCGROWN G., THORNCROFT M., TRICKER K.J., TEARE M.D.,  
 RA SANTIBANEZ-KOREF M.F., HOULSTON R.S., MARTIN J., BIRCH J.M.,  
 RN J. MED. GENET. 32:942-945(1995).  
 [34]  
 RA VARIANTS BA PHE-176; SER-245; TRP-248; TRP-282 AND GLN-286.  
 RX MEDLINE; 96233927.  
 RA AUDREZET M.-P., ROBASZKIEWICZ M., MERCIER B., NOUSBAUM J.-B.,  
 RA HARDY E., BAILL J.-P., VOLANT A., LOZAC'H P., GOUEROU H., FEREC C.;  
 RN HUM. MUTAT. 7:109-113(1996).

Note: remainder of annotations omitted.

Query Match 100.0%; Score 63; DB 1; Length 393;

Best Local Similarity 100.0%; Pred. No. 2.60e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 154 GTRVAMAI 162  
 QY 1 GTRVAMAI 9

RESULT 10  
 ID P53\_MESAU STANDARD; PRT; 396 AA.  
 AC Q00366; P97276.  
 DT 01-DEC-1992 (REL. 24, CREATED)  
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS MESOCRICETUS AURATUS (GOLDEN HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; RODENTIA.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-STRIAN; TISSUE-KIDNEY;  
 RX MEDLINE; 92210007.  
 RA LEGROS Y., MCINTYRE P., SOUSSI T.;  
 RN GENE 112:247-250(1992).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA HOU E.W., WISEMAN R.;  
 RN SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
 TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 BAX AND BAX ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG. NO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL; M75144; G191415; -;  
 DR EMBL; U0182; G473579; -;  
 DR PIR; J06633; J06633.  
 DR HSSP; P04637; IPES.

DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 77  
 FT DOMAIN 78 153  
 FT DOMAIN 319 393  
 FT DOMAIN 314 326  
 FT MOD RES 395 395  
 FT CONFLICT 188 188  
 FT SEQUENCE 396 AA; 43631 MW; C2668ADE CRC32;  
 G -> S (IN REF. 2).

Query Match 100.0%; Score 63; DB 1; Length 396;

Best Local Similarity 100.0%; Pred. No. 2.60e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 157 GTRVAMAI 165  
 QY 1 GTRVAMAI 9

RESULT 11  
 ID P53\_MOUSE STANDARD; PRT; 390 AA.  
 AC P02340;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR TRP53 OR P53.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; RODENTIA.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85027173.  
 RA BIENZ B., ZAKUT-HOURI R., GIOVL D., OREN M.;  
 RN EMO J. 3:2179-2183(1984).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 84068204.  
 RA ZAKUT-HOURI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIOVL D.;  
 RN NATURE 306:594-597(1983).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 84272240.  
 RA JENKINS J.R., RUDGE K., REDMOND S., WADE-EVANS A.;  
 RN NUCLEIC ACIDS RES. 12:5609-5626(1984).  
 [4]  
 RP SEQUENCE FROM N.A. (CLONES PCD53; P53-M1 AND P53-M8).  
 RX MEDLINE; 87064640.  
 RA ARAI N., NOMURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.,  
 RA ROTTEN V.;  
 RN MOL. CELL. BIOL. 6:3232-3239(1986).  
 [5]  
 RP SEQUENCE OF 222-258 FROM N.A.  
 RX MEDLINE; 92115342.  
 RA BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BRENNER R.,  
 RN ONCOGENE 6:2363-2369(1991).  
 [6]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE; 86149247.  
 RA SAMAD A., ANDERSON C.W., CARROLL R.B.;  
 RN PROC. NATL. ACAD. SCI. U.S.A. 83:897-901(1986).  
 [7]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE; 91006019.  
 RA MEK D.W., SIMON S., KIRKAMA U., ECKHART W.;  
 RN EMO J. 9:3253-3260(1990).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A



TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.

-1- SUBCELLULAR LOCATION: NUCLEAR.  
-1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.

-1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

EMBL; X00876; G871421; -  
EMBL; X00877; G871421; JOINED.  
EMBL; X00878; G871421; JOINED.  
EMBL; X00879; G871421; JOINED.  
EMBL; X00880; G871421; JOINED.  
EMBL; X00881; G871421; JOINED.  
EMBL; X00882; G871421; JOINED.  
EMBL; X00883; G871421; JOINED.  
EMBL; X00884; G871421; JOINED.  
EMBL; X00885; G871421; JOINED.  
EMBL; K01700; G200205; -  
EMBL; X01237; G53576; -  
EMBL; X00741; G53571; -  
EMBL; M13872; G200199; -  
EMBL; M13873; G200201; -  
EMBL; M13874; G200203; -  
EMBL; S7930; G243255; -  
PIR; A02684; DNMS53.  
PIR; A22739; A22739.  
PIR; S38822; S38822.  
HSSP; P04637; 1PES.  
DR TRANSFAC; T01806; -  
DR MGD; MGI:98834; TRP53.  
DR PROSITE; PS00348; P53; 1.  
DR ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS; DISEASE MUTATION.  
FT DOMAIN 1 75  
FT 76 150  
FT 276 390  
DOMAIN HIGHLY BASIC AND MAY BE INVOLVED IN  
INTERACTION WITH DNA.  
DOMAIN NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD.RES 312 320  
FT 312 320  
FT MOD.RES 389 389  
FT 135 135  
VARIANT A -> V (CAN COOPERATE WITH AN ACTIVATED  
RAS TO TRANSFORM FIBROBLASTS).  
FT VARIANT 168 168  
FT 168 168  
FT CONFLICT 48 48  
FT CONFLICT 79 81  
FT 81  
SEQUENCE 390 AA; 43458 MW; 8943DD93 CRC32;

Query Match 95.2%; Score 60; DB 1; Length 390;  
Best Local Similarity 88.9%; Pred. No. 1.77e-03;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 151 GSRVAMAI 159  
QY 1 GTRVAMAI 9

RESULT 12  
ID P53 CERAE STANDARD: PRT: 393 AA.  
AC P13481.  
DT 01-JAN-1990 (REL. 13, CREATED)  
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS CERCOPTHECUS AETHIOPS (GREEN MONKEY) (GRIYET)  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.

TISSUE-LIVER;  
MEDLINE: 90045967.  
RA RIGAUDY P., ECKHARDT W.;  
RL NUCLEIC ACIDS RES. 17:8375-8375(1989).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.

-1- SUBCELLULAR LOCATION: NUCLEAR.  
-1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.

-1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

EMBL; X16384; G22796; -  
PIR; S06594; S06594.  
DR HSSP; P04637; 1OLG.  
DR PROSITE; PS00348; P53; 1.  
DR ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 1 68  
FT 68 150  
FT 319 393  
DOMAIN HIGHLY BASIC AND MAY BE INVOLVED IN  
INTERACTION WITH DNA.  
FT DOMAIN 311 323  
FT 311 323  
FT MOD.RES 392 392  
FT 392 392  
SEQUENCE 393 AA; 43696 MW; BB7DC62 CRC32;

Query Match 95.2%; Score 60; DB 1; Length 393;  
Best Local Similarity 88.9%; Pred. No. 1.77e-03;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 154 GSRVAMAI 162  
QY 1 GTRVAMAI 9

RESULT 13  
ID RPOE\_SULAC STANDARD: PRT: 248 AA.  
AC P39466;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE DNA-DIRECTED RNA POLYMERASE SUBUNIT E (EC 2.7.7.6).  
GN RPOE.  
OS SULFOLOBUS ACIDOCALDARIUS.  
OC ARCHAEABACTERIA; CRENARCHAEOTA; SULFOBACLES.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
PC STRAIN-DSM 639.  
RX MEDLINE: 94173739.  
RA LANGER D., LOTTSPEICH F., ZILLIG W.;  
RL NUCLEIC ACIDS RES. 22:694-694(1994).  
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
SUBSTRATES.  
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +  
RNA(N).  
CC -1- SUBUNIT: THE S. ACIDOCALDARIUS RNAP IS COMPOSED OF 13 SUBUNITS.  
CC -1- SIMILARITY: CONTAINS A COPY OF THE 'S1 MOTIF'.  
EMBL; X75411; G415999; -  
EMBL; X38658; S38658.  
PIR; S42389; S42389.  
DR PIR; S42389; S42389.  
DR TRANSCRIPTION; DNA-DIRECTED RNA POLYMERASE; ZINC-FINGER.  
FT ZN\_FING 196 213  
FT 213  
SEQUENCE 248 AA; 27632 MW; AE1B2336 CRC32;

Query Match 82.5%; Score 52; DB 1; Length 248;

Best Local Similarity 77.8%; Pred. No. 2,26e+01;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 139 GDRVRAM11 147

OY 1 GTRVRAMAI 9

RESULT 14

ID P41638; STANDARD; PRT: 198 AA.

AC P41638; 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DE CHLOROPLAST 30S RIBOSOMAL PROTEIN S4.

OS PINUS THUNBERGII (GREEN PINE) (JAPANESE BLACK PINE).

OC CHLOROPLAST.

OC EUKARYOTA; PLANTA; EMBRYOPHYTA; CONIFEROPHYCEAE.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 95024047.

RA WAKASUGI T., TSUDZUKI J., ITO S., NAKASHIMA K., TSUDZUKI T.,

RA SUGIURA M.; PROC. NATL. ACAD. SCI. U.S.A. 91:9794-9798(1994).

CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.

DR EMBL: D17510; G1262697; -

DR PROSITE: PS00632; RIBOSOMAL\_S4; FALSE\_NEG.

KM RIBOSOMAL PROTEIN; CHLOROPLAST.

SO SEQUENCE 198 AA; 23488 MW; 18A580E2 CRC32;

Query Match 77.8%; Score 49; DB 1; Length 198;

Best Local Similarity 66.7%; Pred. No. 1.24e+00; Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 77 GSRVRSIAI 85

OY 1 GTRVRAMAI 9

RESULT 15

ID P53\_FELCA STANDARD; PRT: 386 AA.

AC P41685; 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DE CELLULAR TUMOR ANTIGEN P53.

OS FELIS SILVESTRIIS CATUS (CAT).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; CARNIVORA.

RN [1]

RP SEQUENCE FROM N.A.

RX TISSUE-LYMPH NODE;

RA OKUDA M., UEDA A., SAKAI T., OHASHI T., MOMOI Y., YOUN H.Y.,

RA WATARI T., GOITSUKA R., TSUJIMOTO H., HASEGAWA A.;

INT. J. CANCER 58:602-607(1994).

RN [2]

RP SEQUENCE OF 34-354 FROM N.A.

RX MEDLINE: 94114699.

RA OKUDA M., UEDA A., MATSUMOTO Y., MOMOI Y., WATARI T., GOITSUKA R.,

RA O'BRIEN S.J., TSUJIMOTO H., HASEGAWA A.;

INT. J. VET. MED. SCI. 55:801-805(1993).

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES

GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL

CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN

TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A

TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION

BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF

THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.

APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF

BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2

CC EXPRESSION.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY

OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED

IN MANY TYPES OF CANCER.

CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

DR EMBL: D26608; G538225; -

DR EMBL: D16460; G575528; -

DR PROSITE: PS00348; P53; 1.

KM ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;

KM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.

FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).

FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).

FT CONFLICT 285 285 K -> R (IN REF. 2).

SO SEQUENCE 386 AA; 42692 MW; D6C7132A CRC32;

Query Match 77.8%; Score 49; DB 1; Length 386;

Best Local Similarity 88.9%; Pred. No. 1.24e+00;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 146 GTCVRAMAI 154

OY 1 GTRVRAMAI 9

Search completed: Fri Sep 11 13:37:21 1998

Job time : 6 secs.

# MUSE RELEASE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

Msrch: pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:37:38 1998; Maspar time 3.91 Seconds

Tabular output not generated. 97.020 Million cell updates/sec

Title: >US-08-452-843-16  
Description: (1-9) from US08452843.pep  
Perfect Score: 63

Sequence: 1 GTRVAMAI 9

Scoring table: PAM 150  
Gap 15

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

split6  
1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mhc 6:sp\_organelle 7:sp\_phase 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp\_vertebrate  
13:sp\_unclassified

Statistics: Mean 22.343; Variance 24.231; scale 0.922

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description            | Pred. No. |
|------------|-------|-------------|--------|-----------|------------------------|-----------|
| 1          | 63    | 100.0       | 136    | 10 060434 | TUMOR SUPPRESSOR P53 ( | 6.41e-04  |
| 2          | 63    | 100.0       | 136    | 10 064396 | CELLULAR TUMOR ANTIGEN | 6.41e-04  |
| 3          | 63    | 100.0       | 196    | 4 029484  | CELLULAR TUMOR ANTIGEN | 6.41e-04  |
| 4          | 63    | 100.0       | 205    | 10 035873 | CELLULAR TUMOR ANTIGEN | 6.41e-04  |
| 5          | 63    | 100.0       | 245    | 2 015085  | P53 TRANSFORMATION SUP | 6.41e-04  |
| 6          | 63    | 100.0       | 391    | 13 036006 | CELLULAR TUMOR ANTIGEN | 6.41e-04  |
| 7          | 63    | 100.0       | 393    | 2 015088  | P53 TRANSFORMATION SUP | 6.41e-04  |
| 8          | 63    | 100.0       | 393    | 2 016535  | P53 TRANSFORMATION SUP | 6.41e-04  |
| 9          | 63    | 100.0       | 393    | 2 016810  | CELLULAR TUMOR ANTIGEN | 6.41e-04  |
| 10         | 63    | 100.0       | 393    | 2 015087  | P53 TRANSFORMATION SUP | 6.41e-04  |
| 11         | 63    | 100.0       | 393    | 2 016811  | CELLULAR TUMOR ANTIGEN | 6.41e-04  |
| 12         | 63    | 100.0       | 393    | 2 016809  | CELLULAR TUMOR ANTIGEN | 6.41e-04  |
| 13         | 63    | 100.0       | 393    | 2 016808  | CELLULAR TUMOR ANTIGEN | 6.41e-04  |
| 14         | 63    | 100.0       | 393    | 2 016848  | CELLULAR TUMOR ANTIGEN | 6.41e-04  |
| 15         | 63    | 100.0       | 393    | 2 016807  | CELLULAR TUMOR ANTIGEN | 6.41e-04  |
| 16         | 63    | 100.0       | 393    | 2 015086  | P53 TRANSFORMATION SUP | 6.41e-04  |
| 17         | 60    | 95.2        | 90     | 10 P70656 | P53 (FRAGMENT)         | 4.01e-03  |
| 18         | 60    | 95.2        | 135    | 10 064451 | CELLULAR TUMOR ANTIGEN | 4.01e-03  |
| 19         | 60    | 95.2        | 238    | 11 P89004 | P53 (FRAGMENT)         | 4.01e-03  |
| 20         | 60    | 95.2        | 286    | 11 P89003 | P53 (FRAGMENT)         | 4.01e-03  |

|    |    |      |      |           |                        |          |
|----|----|------|------|-----------|------------------------|----------|
| 21 | 60 | 95.2 | 286  | 11 P90332 | P53 (FRAGMENT)         | 4.01e-03 |
| 22 | 60 | 95.2 | 378  | 11 P89002 | P53 (FRAGMENT)         | 4.01e-03 |
| 23 | 54 | 85.7 | 364  | 9 026392  | SENSORY TRANSDUCTION H | 1.36e-01 |
| 24 | 53 | 84.1 | 567  | 9 026456  | SENSORY TRANSDUCTION H | 2.40e-01 |
| 25 | 50 | 79.4 | 386  | 7 038068  | TAIL SHEATH PROTEIN (G | 1.26e+00 |
| 26 | 49 | 77.8 | 495  | 9 026559  | SENSORY TRANSDUCTION H | 2.17e+00 |
| 27 | 48 | 76.2 | 209  | 9 P71647  | HYPOTHETICAL 22.5 KD P | 3.69e+00 |
| 28 | 48 | 76.2 | 383  | 9 007407  | ALCOHOL DEHYDROGENASE  | 3.69e+00 |
| 29 | 47 | 74.6 | 172  | 9 P94384  | DNA FOR 25-36 DEGREE R | 6.24e+00 |
| 30 | 47 | 74.6 | 244  | 9 031473  | YCGI PROTEIN           | 6.24e+00 |
| 31 | 47 | 74.6 | 602  | 1 018739  | ECOCITLINASE           | 6.24e+00 |
| 32 | 47 | 74.6 | 1209 | 3 P91581  | EARLY LIGHT-INDUCED PR | 1.05e+01 |
| 33 | 46 | 73.0 | 192  | 8 P93169  | FERRDOXIN REDUCTASE O  | 1.05e+01 |
| 34 | 46 | 73.0 | 411  | 9 051747  | ISOPROPYLBENZENE-2,3-D | 1.05e+01 |
| 35 | 46 | 73.0 | 411  | 9 P95369  | SIGNAL TRANSDUCER AND  | 1.05e+01 |
| 36 | 46 | 73.0 | 748  | 2 004765  | HOMEOBOX PROTEIN HX1 ( | 1.74e+01 |
| 37 | 45 | 71.4 | 94   | 3 005010  | RIBULOSE BIPHOSPHATE   | 1.74e+01 |
| 38 | 45 | 71.4 | 171  | 8 042900  | DNAJ-HOMOLOGUE         | 1.74e+01 |
| 39 | 45 | 71.4 | 280  | 9 P77642  | COSE7.2                | 1.74e+01 |
| 40 | 45 | 71.4 | 286  | 3 017670  | HYPOHETICAL 35.9 KD P  | 1.74e+01 |
| 41 | 45 | 71.4 | 324  | 1 013780  | GCM PROTEIN            | 1.74e+01 |
| 42 | 45 | 71.4 | 504  | 3 027403  | DNA GYRASE SUBUNIT A ( | 1.74e+01 |
| 43 | 45 | 71.4 | 907  | 9 059784  | TSC2 PROTEIN (FRAGMENT | 1.74e+01 |
| 44 | 45 | 71.4 | 1782 | 12 042180 | L PROTEIN              | 1.74e+01 |
| 45 | 45 | 71.4 | 4036 | 11 066431 |                        |          |

## ALIGNMENTS

|                       |   |   |              |         |         |
|-----------------------|---|---|--------------|---------|---------|
| ID                    | 060434  | 1                                       | PRELIMINARY; | PRT;    | 136 AA. |
| AC                    | 060434;   | P97257;                                 |              |         |         |
| DT                    | 01-NOV-1986   | (TREMBLREL. 01, CREATED)                |              |         |         |
| DT                    | 01-JAN-1998   | (TREMBLREL. 03, LAST SEQUENCE UPDATE)   |              |         |         |
| DT                    | 01-JAN-1998   | (TREMBLREL. 05, LAST ANNOTATION UPDATE) |              |         |         |
| DE                    | TUMOR SUPPRESSOR P53 (FRAGMENT).                                  |   |              |         |         |
| OS                    | CRICETULUS GRISEUS (CHINESE HAMSTER).                             |   |              |         |         |
| OC                    | EURAROTIA, METAEOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.    |   |              |         |         |
| CC                    | EUTHERIA; RODENTIA.   |   |              |         |         |
| RN                    | [1]   |   |              |         |         |
| RP                    | SEQUENCE FROM N.A.  |   |              |         |         |
| RA                    | MAI S., FLURI M., SIMARSKI D., HUPPI K.;                          |   |              |         |         |
| RL                    | SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.              |   |              |         |         |
| CC                    | -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT |   |              |         |         |
| CC                    | PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL   |   |              |         |         |
| CC                    | CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY |   |              |         |         |
| CC                    | REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED |   |              |         |         |
| CC                    | FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF   |   |              |         |         |
| CC                    | CYCLIN-DEPENDENT KINASES (BY SIMILARITY).                         |   |              |         |         |
| CC                    | -1- SUBCELLULAR LOCATION: NUCLEAR.                                |   |              |         |         |
| DR                    | EMBL; U41452; G1786177; "-  |   |              |         |         |
| DR                    | PROSITE; PS00348; P53; 1.   |   |              |         |         |
| KW                    | ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  |   |              |         |         |
| KW                    | NUCLEAR PROTEIN; PHOSPHORYLATION.                                 |   |              |         |         |
| FT                    | NON_TER   | 1                                       |              |         |         |
| FT                    | NON_TER   | 136                                     |              |         |         |
| SO                    | SEQUENCE  | 136 AA; 15438 MW; 10679AD4 CRC32;       |              |         |         |
| Query Match           |   | 100.0%; Score 63; DB 10; Length 136;    |              |         |         |
| Best Local Similarity |   | 100.0%; Pred. No. 6.41e-04;             |              |         |         |
| Matches               | 9; Conservative   | 0; Mismatches                           | 0; Indels    | 0; Gaps |         |
| DB                    | 23 GTRVAMAI 31  |   |              |         |         |
|                       |   |   |              |         |         |
| QY                    | 1 GTRVAMAI 9  |   |              |         |         |
| RESULT                | 2   | PRELIMINARY;                            | PRT;         | 136 AA. |         |
| ID                    | 064396  |   |              |         |         |
| AC                    | 064396;   | P97940;                                 |              |         |         |
| DT                    | 01-NOV-1996   | (TREMBLREL. 01, CREATED)                |              |         |         |
| DT                    | 01-JAN-1998   | (TREMBLREL. 05, LAST SEQUENCE UPDATE)   |              |         |         |
| DT                    | 01-JAN-1998   | (TREMBLREL. 05, LAST ANNOTATION UPDATE) |              |         |         |
| DT                    | 01-JAN-1998   | (TREMBLREL. 05, LAST SEQUENCE UPDATE)   |              |         |         |

DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS CRICENTULUS GRISEUS (CHINESE HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MAI S., FLURI M., SIMWASKI D., HUPPI K.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; U04151; G176175; -.  
 DR PROSITE; PS00348; P53; 1.  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ANTI-ONCOGENE; DNA-BINDING;  
 KW TRANSCRIPTION REGULATION; ACTIVATOR.  
 FT NON TER 1 1  
 SO SEQUENCE 136 AA; 13411 MW; CEB916C9 CRC32;  
 Query Match 100.0%; Score 63; DB 10; Length 136;  
 Best Local Similarity 100.0%; Pred. No. 6,41e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 23 GTRVRAAI 31  
 Oy 1 GTRVRAAI 9  
 RESULT 3  
 ID 029484 PRELIMINARY; PRT; 196 AA.  
 AC 029484;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS EQUUS CABALLUS (HORSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; PERISSODACTYLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BOCHER K., SZALAI G., MARTI E., PAULI U., LAZARY S.;  
 RL RES. VET. SCI. 0:0-0(0).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; X91793; E218035; -.  
 DR PROSITE; PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT NON TER 1 1  
 SO SEQUENCE 196 AA; 22080 MW; F443239C CRC32;  
 Query Match 100.0%; Score 63; DB 4; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 6,41e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 20 GTRVRAAI 28  
 Oy 1 GTRVRAAI 9  
 RESULT 4

ID 035873 PRELIMINARY; PRT; 205 AA.  
 AC 035873;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS CRICENTULUS GRISEUS (CHINESE HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA RAINALDI G., MARCHETTI S., CAPECCHI B., MENEVERI R., PIRAS A.,  
 RA LEZZI R.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; U74487; G2581764; -.  
 DR PROSITE; PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT NON TER 1 1  
 SO SEQUENCE 205 AA; 23122 MW; 680DDDC CRC32;  
 Query Match 100.0%; Score 63; DB 10; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 6,41e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 29 GTRVRAAI 37  
 Oy 1 GTRVRAAI 9  
 RESULT 5  
 ID 015085 PRELIMINARY; PRT; 245 AA.  
 AC 015085;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., YOUSSEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL; X60010; G506433; -.  
 FT NON TER 245 245  
 SO SEQUENCE 245 AA; 27066 MW; 55B80C07 CRC32;  
 Query Match 100.0%; Score 63; DB 2; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 6,41e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 154 GTRVRAAI 162  
 Oy 1 GTRVRAAI 9  
 RESULT 6

```

ID 036006 PRELIMINARY: PRT: 391 AA.
AC 036006;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53.
GN P53.
OS MARMATA MONAX.
OG PLASMID PT7BLUE (R).
OC UNCLASSIFIED.
RN [1]
RP SEQUENCE FROM N.A.
RA FEITELSON M.A., RANGANATHAN P.N., CLAYTON M.M., ZHANG S.M.;
ONCOGENE 15:327-336(1997).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL, AJ001022; E31287; -.
DR PROSITE; PS00348; P53; 1.
KW PLASMID; ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION;
KW ACTIVATOR; NUCLEAR PROTEIN; PHOSPHORYLATION.
SQ SEQUENCE 391 AA; 43468 MW; 95FABBF2 CRC32;

Query Match
Best Local Similarity 100.0%; Score 63; DB 13; Length 391;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 152 GTRVAMAI 160
OY 1 GTRVAMAI 9

RESULT 7 PRELIMINARY: PRT: 393 AA.
ID 015088
AC 015088;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN P53.
OS HOMO SAPIENS (HUMAN).
OC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
EMBO J. 10:2879-2887(1991).
RL EMBL; X60016; G506445; -.
FT VARIANT 238 238 Y -> C.
FT NON TER 393 393
SQ SEQUENCE 393 AA; 43713 MW; A01E1523 CRC32;

Query Match
Best Local Similarity 100.0%; Score 63; DB 2; Length 393;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVAMAI 162
OY 1 GTRVAMAI 9

RESULT 8 PRELIMINARY: PRT: 393 AA.
ID 016535
AC 016535;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

```

```

GN P53.
OS HOMO SAPIENS (HUMAN).
OC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
EMBO J. 10:2879-2887(1991).
RL EMBL; X60017; G506447; -.
DR EMBL; X60015; G506443; -.
FT VARIANT 248 248 Q -> R.
FT NON TER 393 393
SQ SEQUENCE 393 AA; 43684 MW; 239818A9 CRC32;

Query Match
Best Local Similarity 100.0%; Score 63; DB 2; Length 393;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVAMAI 162
OY 1 GTRVAMAI 9

RESULT 9 PRELIMINARY: PRT: 393 AA.
ID 016810
AC 016810;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS HOMO SAPIENS (HUMAN).
OC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
EMBO J. 10:2879-2887(1991).
RL EMBL; X60020; G506453; -.
DR EMBL; X60020; G506453; -.
DR PROSITE; PS00348; P53; 1.
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW NUCLEAR PROTEIN; PHOSPHORYLATION.
FT VARIANT 254 254 D -> N.
FT VARIANT 254 254 D -> V.
FT NON TER 393 393
SQ SEQUENCE 393 AA; 43714 MW; 5F914579 CRC32;

Query Match
Best Local Similarity 100.0%; Score 63; DB 2; Length 393;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVAMAI 162
OY 1 GTRVAMAI 9

RESULT 10 PRELIMINARY: PRT: 393 AA.
ID 015087
AC 015087;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

```

OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RL EMO J. 10:2879-2887(1991).  
DR EMBL: X60014: G506441: -  
FT VARIANT 237 237 I -> M.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA: 43694 MM: 9BB81992 CRC32;  
Query Match 100.0%; Score 63; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 6.41e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 154 GTRVRAAI 162  
QY 1 GTRVRAAI 9  
RESULT 11  
ID 016811 PRELIMINARY; PRT; 393 AA.  
AC 016811;  
DT 01-NOV-1996 (TREMBLER. 01, CREATED)  
DT 01-NOV-1996 (TREMBLER. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLER. 05, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 8512634.  
RA MATLASHWSTI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.;  
RL BENCHIMOL S.;  
RL EMO J. 3:3257-3262(1984).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87064416.  
RA LAMB P., CRAWFORD L.;  
RL MOL. CELL. BIOL. 6:1379-1385(1986).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC EMBL: M13121: G386994: -  
DR EMBL: M13112: G386994: JOINED.  
DR EMBL: M13113: G386994: JOINED.  
DR EMBL: M13114: G386994: JOINED.  
DR EMBL: M13115: G386994: JOINED.  
DR EMBL: M13116: G386994: JOINED.  
DR EMBL: M13117: G386994: JOINED.  
DR EMBL: M13118: G386994: JOINED.  
DR EMBL: M13119: G386994: JOINED.  
DR EMBL: M13120: G386994: JOINED.  
DR PROSITE: PS00348: P53; 1.  
KW REPEAT: TUMOR ANTIGEN; ANTI-ONCOGENE; DNA-BINDING;  
KW TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN;  
KW PHOSPHORYLATION.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA: 43698 MM: 3EA71431 CRC32;  
Query Match 100.0%; Score 63; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 6.41e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 154 GTRVRAAI 162  
QY 1 GTRVRAAI 9

QY 1 GTRVRAAI 9  
RESULT 12  
ID 016809 PRELIMINARY; PRT; 393 AA.  
AC 016809;  
DT 01-NOV-1996 (TREMBLER. 01, CREATED)  
DT 01-NOV-1996 (TREMBLER. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLER. 05, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RL EMO J. 10:2879-2887(1991).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC EMBL: X60019: G506451: -  
DR PROSITE: PS00348: P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
FT VARIANT 213 213 Q -> R.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA: 43684 MM: CB70BD7F CRC32;  
Query Match 100.0%; Score 63; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 6.41e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 154 GTRVRAAI 162  
QY 1 GTRVRAAI 9  
RESULT 13  
ID 016808 PRELIMINARY; PRT; 393 AA.  
AC 016808;  
DT 01-NOV-1996 (TREMBLER. 01, CREATED)  
DT 01-NOV-1996 (TREMBLER. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLER. 05, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RL EMO J. 10:2879-2887(1991).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC EMBL: X60018: G506449: -  
DR PROSITE: PS00348: P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
FT VARIANT 163 163 H -> Y.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA: 43627 MM: AFD8A9E3 CRC32;

Query Match 100.0%; Score 63; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.41e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 154 GTRVAMAI 162  
 |||||||  
 OY 1 GTRVAMAI 9

RESULT 14  
 ID Q16848 PRELIMINARY; PRT; 393 AA.

AC Q16848;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.

OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTER V.;

RL MOL. CELL. BIOL. 6:4650-4656(1986).  
 CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL: M14694; G339814; -.  
 DR PROSITE: PS00348; P53; 1.  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ANTI-ONCOGENE; DNA-BINDING;  
 KM TRANSCRIPTION REGULATION; ACTIVATOR.  
 SQ SEQUENCE 393 AA; 43723 MW; DA7D302F CRC32;

Query Match 100.0%; Score 63; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.41e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 154 GTRVAMAI 162  
 |||||||  
 OY 1 GTRVAMAI 9

RESULT 15  
 ID Q16807 PRELIMINARY; PRT; 393 AA.

AC Q16807;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.

OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;

RL EMBL J. 10:2879-2887(1991).  
 CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: X60011; G506435; -.

DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 193 193 R -> H.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43731 MW; 279BC9CB CRC32;

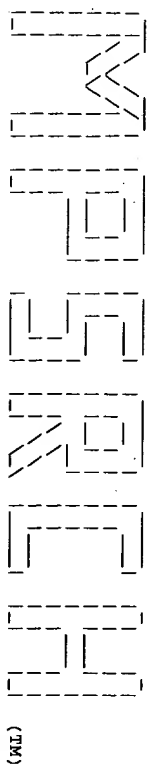
Query Match 100.0%; Score 63; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.41e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 154 GTRVAMAI 162  
 |||||||  
 OY 1 GTRVAMAI 9

Search completed: Fri Sep 11 13:38:19 1998  
 Job time : 41 secs.

**THIS PAGE BLANK (USPTO)**





(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mparch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:32:14 1998; Maspar time 2.63 Seconds

Tabular output not generated. 67.677 Million cell updates/sec

Title: >US-08-452-843-15  
Description: (1-11) from US08452843.pep  
Perfect Score: 86  
Sequence: 1 SPALNMFCOL 11

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

a-geneseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 17.801; Variance 50.624; scale 0.352

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID  | Description           | Pred. No. |
|------------|-------|-------------|--------|--------|-----------------------|-----------|
| 1          | 86    | 100.0       | 28 20  | W03363 | p53 protein residues  | 3.80e-03  |
| 2          | 86    | 100.0       | 253 24 | W28484 | Human p53 protein var | 3.80e-03  |
| 3          | 86    | 100.0       | 253 24 | W28483 | Human p53 protein var | 3.80e-03  |
| 4          | 86    | 100.0       | 270 24 | W28486 | Human p53 protein var | 3.80e-03  |
| 5          | 86    | 100.0       | 270 24 | W28485 | Human p53 protein var | 3.80e-03  |
| 6          | 86    | 100.0       | 319 24 | W28495 | Human p53 protein var | 3.80e-03  |
| 7          | 86    | 100.0       | 319 24 | W28496 | Human p53 protein var | 3.80e-03  |
| 8          | 86    | 100.0       | 335 24 | W28497 | Human p53 protein var | 3.80e-03  |
| 9          | 86    | 100.0       | 335 24 | W28498 | Human p53 protein var | 3.80e-03  |
| 10         | 86    | 100.0       | 353 24 | W28494 | Human p53 protein var | 3.80e-03  |
| 11         | 86    | 100.0       | 353 24 | W28493 | Human p53 protein var | 3.80e-03  |
| 12         | 86    | 100.0       | 363 21 | W13971 | Modified p53 variant  | 3.80e-03  |
| 13         | 86    | 100.0       | 363 21 | W13974 | Modified p53 variant  | 3.80e-03  |
| 14         | 86    | 100.0       | 363 21 | W13973 | Modified p53 variant  | 3.80e-03  |
| 15         | 86    | 100.0       | 363 21 | W13972 | Modified p53 variant  | 3.80e-03  |
| 16         | 86    | 100.0       | 363 24 | W28479 | Human p53 protein var | 3.80e-03  |
| 17         | 86    | 100.0       | 363 24 | W28480 | Human p53 protein var | 3.80e-03  |
| 18         | 86    | 100.0       | 363 21 | W13975 | Modified p53 variant  | 3.80e-03  |

|    |    |       |        |        |                       |          |
|----|----|-------|--------|--------|-----------------------|----------|
| 19 | 86 | 100.0 | 363 21 | W13977 | Modified p53 variant  | 3.80e-03 |
| 20 | 86 | 100.0 | 370 21 | W13977 | Chimeric p53 protein. | 3.80e-03 |
| 21 | 86 | 100.0 | 374 24 | W28481 | Human p53 protein var | 3.80e-03 |
| 22 | 86 | 100.0 | 374 24 | W28482 | Human p53 protein var | 3.80e-03 |
| 23 | 86 | 100.0 | 381 24 | W28490 | Human p53 protein var | 3.80e-03 |
| 24 | 86 | 100.0 | 381 24 | W28489 | Human p53 protein var | 3.80e-03 |
| 25 | 86 | 100.0 | 393 24 | W25155 | Human p53 variant fou | 3.80e-03 |
| 26 | 86 | 100.0 | 393 22 | W13953 | T284K modified human  | 3.80e-03 |
| 27 | 86 | 100.0 | 393 22 | W13948 | Human wild-type p53 t | 3.80e-03 |
| 28 | 86 | 100.0 | 393 21 | W05345 | Human p53 mutant N239 | 3.80e-03 |
| 29 | 86 | 100.0 | 393 22 | W13951 | Human tumour-derived  | 3.80e-03 |
| 30 | 86 | 100.0 | 393 22 | W13979 | Human tumour-derived  | 3.80e-03 |
| 31 | 86 | 100.0 | 393 22 | W13952 | Human tumour-derived  | 3.80e-03 |
| 32 | 86 | 100.0 | 393 18 | R91933 | Wild type p53 protein | 3.80e-03 |
| 33 | 86 | 100.0 | 393 21 | W05348 | Human p53 mutant R282 | 3.80e-03 |
| 34 | 86 | 100.0 | 393 21 | W05344 | Human p53             | 3.80e-03 |
| 35 | 86 | 100.0 | 393 21 | W13968 | Modified p53 variant  | 3.80e-03 |
| 36 | 86 | 100.0 | 393 21 | W13970 | Modified p53 variant  | 3.80e-03 |
| 37 | 86 | 100.0 | 393 21 | W13969 | Modified p53 variant  | 3.80e-03 |
| 38 | 86 | 100.0 | 401 24 | W28487 | Human p53 protein var | 3.80e-03 |
| 39 | 86 | 100.0 | 401 24 | W28488 | Human p53 protein var | 3.80e-03 |
| 40 | 86 | 100.0 | 406 21 | W13964 | Chimeric p53 protein. | 3.80e-03 |
| 41 | 86 | 100.0 | 406 21 | W13966 | Chimeric p53 protein. | 3.80e-03 |
| 42 | 86 | 100.0 | 411 21 | W13967 | Chimeric p53 protein. | 3.80e-03 |
| 43 | 86 | 100.0 | 533 23 | W19763 | p53-GM-CSF immunostim | 3.80e-03 |
| 44 | 86 | 100.0 | 535 24 | W28492 | Human p53 protein var | 3.80e-03 |
| 45 | 86 | 100.0 | 535 24 | W28491 | Human p53 protein var | 3.80e-03 |

## ALIGNMENTS

RESULT 1  
ID W03363 standard; peptide; 28 AA.  
AC W03363;  
DT 10-MAR-1997 (first entry)  
DE p53 protein residues 124-151.  
KW Cytotoxic T lymphocyte; CTL; epitope; p53; template;  
KM Ratchet library; pharmaceutical; vaccine; treatment; prevention;  
KN disease; malignancy; cancer.  
OS Homo sapiens.  
FH Key  
FT region 11..20 Location/Qualifiers  
FT /note="cytotoxic T lymphocyte epitope"  
FN W09622067-A2.  
PD 25-JUL-1996.  
PR 15-DEC-1995; U16290.  
PR 27-DEC-1994; US-366332.  
PA (UNBL-) UNITED BIOMEDICAL INC.  
PI Kuebler PJ, Nixon DF;  
DR WPI: 96-354273/35.  
PT Ratchet library of peptide(s) contg. an immuno:stimulatory CTL  
PT epitope - derived from longer template peptide, useful as  
PT pharmaceutical or vaccine against infectious disease or malignancy  
PS Claim 9: Pages 36-37; 60pp; English.  
CC The present peptide comprises residues 124-151 of the p53 protein,  
CC contains cytotoxic T lymphocyte (CTL) epitope and can be used as a  
CC template in the prepn. of a ratchet library, comprising peptides  
CC contg. at least 1 immunostimulatory CTL epitope. Basically the  
CC distribution of amino acids at each position in the template is  
CC calculated, a ratchet library constructed from the longer template  
CC peptide by sequentially ratcheting it into the shorter ratchet  
CC length and the peptides synthesised using standard solid phase  
CC methods. The library can be used in pharmaceuticals and vaccines  
CC for the treatment and/or prevention of diseases and malignancies  
CC associated with p53 mutation, e.g. cancer.  
CC Several epitopes can be incorporated into the same library, rather  
CC than using a mixt. of individually synthesised immunogenic  
CC peptides, which helps to overcome problems of genetic diversity  
CC and MHC restriction. The library may also include antigenic  
CC variations and escape mutations.  
SQ Sequence 28 AA:  
Query Match 100.0%; Score 86; DB 20; Length 28;

Best Local Similarity 100.0%; Pred. No. 3.80e-03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 4 spalnkmfcq1 14  
| | | | | | | | | |  
OY 1 SPALNKMFCOL 11

RESULT 2  
ID W28484 standard; Protein: 253 AA.  
AC W28484;

DE 25-NOV-1997. (first entry)  
KW Human p53 protein; variant V-367H.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten; transactivation domain (TD) from herpes simplex virus viral protein VP16 (amino acids 411-490). The present sequence is that of a specifically claimed p53 variant designated V-367 and comprising the VP16 TD with amino acids 75-367 of human wild-type p53 (but with Arg182 replaced by His). The p53 variants are more active and more stable tumour suppressors and apoptosis-inducing agents than wild-type p53 and are active where the wild-type protein is not.  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Chimeric - Homo sapiens.  
OS Chimeric - Herpes simplex virus.  
OS Synthetic.

FT Key Location/Qualifiers  
FT misc\_difference 189 /note- "Arg residue at position 182 of wild-type p53 has been mutated to His"

FT W09704092-A1.

PD 06-FEB-1997.

PF 17-JUL-1996; F01111.

PR 19-JUL-1995; FR-008729.

PA (RHON ) RHONE POULENC RORER SA.

PI Bracco L, Conseiller E;

DR WPI; 97-132633/12.

PT New p53 variants e.g. with oligomerisation domain replaced by leucine zipper - useful for treating hyper-proliferative disorders,

PT esp. cancer and restenosis

PS Claim 32; Page -; 133pp; French.

CC Claimed variants of protein p53 have at least part of the p53 transactivation domain (amino acids 1-74) deleted and replaced by the transactivating domain (TD) from herpes simplex virus viral protein VP16 (amino acids 411-490). The present sequence is that of a specifically claimed p53 variant designated V-367 and comprising the VP16 TD and amino acids 75-367 of human wild-type p53 (but with Arg182 replaced by His). The p53 variants are more active and more stable tumour suppressors and apoptosis-inducing agents than wild-type p53 and are active where the wild-type protein is not.

CC (Note: this sequence does not appear in the specification and has been produced by modifying the given sequence of variant V-367).

CC Sequence 253 AA;

CC Query Match

Best Local Similarity 100.0%; Score 86; DB 24; Length 253;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 135 spalnkmfcq1 145  
| | | | | | | | | |  
OY 1 SPALNKMFCOL 11

RESULT 3  
ID W28483 standard; Protein: 253 AA.  
AC W28483;

DE 25-NOV-1997 (first entry)

KW Human p53 protein variant V-367 encoded by p53L141.

KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten; transactivation domain (TD) from herpes simplex virus viral protein VP16 (amino acids 411-490). The present sequence is that of a specifically claimed p53 variant designated V-367 and comprising the VP16 TD with amino acids 75-367 of human wild-type p53 (but with Arg182 replaced by His). The p53 variants are more active and more stable tumour suppressors and apoptosis-inducing agents than wild-type p53 and are active where the wild-type protein is not.

KW anti-oncogene; hyperproliferation; cancer; restenosis;

KW tumour suppression; apoptosis.

OS Chimeric - Homo sapiens.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

PN W09704092-A1.

PD 06-FEB-1997.

PF 17-JUL-1996; F01111.

PR 19-JUL-1995; FR-008729.

PA (RHON ) RHONE POULENC RORER SA.

PI Bracco L, Conseiller E;

DR WPI; 97-132633/12.

PT New p53 variants e.g. with oligomerisation domain replaced by leucine zipper - useful for treating hyper-proliferative disorders,

PT esp. cancer and restenosis

PS Claim 33; Page -; 133pp; French.

CC Claimed variants of protein p53 have at least part of the p53 transactivation domain (amino acids 1-74) deleted and replaced by the transactivating domain (TD) from herpes simplex virus viral protein VP16 (amino acids 411-490). The present sequence is that of a specifically claimed p53 variant designated V-367 and comprising the VP16 TD with amino acids 75-367 of human wild-type p53 (but with Arg182 replaced by His), followed by the last 19 C-terminal amino acids of the alternatively spliced (AS) form of murine p53 (encoded by a synthetic linker). The variants are more active and more stable tumour suppressors and apoptosis-inducing agents than wild-type p53 and are active where the wild-type protein is not.

CC (Note: this sequence does not appear in the specification and has been produced by modifying the given sequence of variant V-AS).

CC Sequence 270 AA;

CC Query Match

Best Local Similarity 100.0%; Score 86; DB 24; Length 270;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PA (RHON ) RHONE POULENC RORER SA.

PI Bracco L, Conseiller E;

DR WPI; 97-132633/12.

DR N-PSDB; T86217.

PT New p53 variants e.g. with oligomerisation domain replaced by leucine zipper - useful for treating hyper-proliferative disorders,

PT esp. cancer and restenosis

PS Claim 32; Pages 80-81; 133pp; French.

CC Claimed variants of protein p53 have at least part of the p53 transactivation domain (amino acids 1-74) deleted and replaced by the transactivating domain (TD) from herpes simplex virus viral protein VP16 (amino acids 411-490). The present sequence is that of a specifically claimed p53 variant designated V-367 and comprising the VP16 TD with amino acids 75-367 of human wild-type p53. The p53 variants are more active and more stable tumour suppressors and apoptosis-inducing agents than wild-type p53 and are active where the wild-type protein is not.

CC Sequence 253 AA;

CC Query Match

Best Local Similarity 100.0%; Score 86; DB 24; Length 253;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 135 spalnkmfcq1 145  
| | | | | | | | | |  
OY 1 SPALNKMFCOL 11

RESULT 4  
ID W28486 standard; Protein: 270 AA.  
AC W28486;

DE 25-NOV-1997 (first entry)

KW Human p53 protein variant V-AS.

KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten; transactivation domain (TD) from herpes simplex virus viral protein VP16 (amino acids 411-490). The present sequence is that of a specifically claimed p53 variant designated V-367 and comprising the VP16 TD with amino acids 75-367 of human wild-type p53 (but with Arg182 replaced by His). The p53 variants are more active and more stable tumour suppressors and apoptosis-inducing agents than wild-type p53 and are active where the wild-type protein is not.

KW anti-oncogene; hyperproliferation; cancer; restenosis; murine; tumour suppression; apoptosis; alternative splicing; AS form.

OS Chimeric - Homo sapiens.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

FT Key Location/Qualifiers

FT misc\_difference 189 /note- "Arg residue at position 182 of wild-type p53 has been mutated to His"

FT W09704092-A1.

PD 06-FEB-1997.

PF 17-JUL-1996; F01111.

PR 19-JUL-1995; FR-008729.

PA (RHON ) RHONE POULENC RORER SA.

PI Bracco L, Conseiller E;

DR WPI; 97-132633/12.

PT New p53 variants e.g. with oligomerisation domain replaced by leucine zipper - useful for treating hyper-proliferative disorders,

PT esp. cancer and restenosis

PS Claim 33; Page -; 133pp; French.

CC Claimed variants of protein p53 have at least part of the p53 transactivation domain (amino acids 1-74) deleted and replaced by the transactivating domain (TD) from herpes simplex virus viral protein VP16 (amino acids 411-490). The present sequence is that of a specifically claimed p53 variant designated V-AS and comprising the VP16 TD with amino acids 75-367 of human wild-type p53 (but with Arg182 replaced by His), followed by the last 19 C-terminal amino acids of the alternatively spliced (AS) form of murine p53 (encoded by a synthetic linker). The variants are more active and more stable tumour suppressors and apoptosis-inducing agents than wild-type p53 and are active where the wild-type protein is not.

CC (Note: this sequence does not appear in the specification and has been produced by modifying the given sequence of variant V-AS).

CC Sequence 270 AA;

CC Query Match

Best Local Similarity 100.0%; Score 86; DB 24; Length 270;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 135 spalnkmfcq1 145  
 1 SPALNKMFCOL 11

RESULT 5  
 ID W28495 standard; Protein: 270 AA.  
 AC W28495:  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-AS encoded by PEC143.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis; murine;  
 KW tumour suppression; apoptosis; alternative splicing; AS form.  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1995: F01111.  
 PR 19-JUL-1995: FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB: T86218.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 33: Pages 82-83; 133pp: French.  
 CC Claimed variants of protein p53 have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TAD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-AS and comprising  
 CC the VP16 TP with amino acids 75-366 of human wild-type p53, followed  
 CC by the last 19 C-terminal amino acids of the alternatively spliced  
 CC (AS) form of murine p53 (encoded by a synthetic linker). The  
 CC variants are more active and more stable tumour suppressors and  
 CC apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not.  
 SQ Sequence 270 AA:

Query Match 100.0%: Score 86; DB 24: Length 270;  
 Best Local Similarity 100.0%: Pred. No. 3.80e-03;  
 Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 135 spalnkmfcq1 145  
 1 SPALNKMFCOL 11

RESULT 6  
 ID W28495 standard; Protein: 319 AA.  
 AC W28495:  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360-325 encoded by PEC178.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1995: F01111.  
 PR 19-JUL-1995: FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB: T86223.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,

PT esp. cancer and restenosis  
 PS Claim 38; Pages 92-94; 133pp: French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360-325 and comprising  
 CC the 325-360 domain, amino acids 75-325 of human wild-type p53 and a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 319 AA:

Query Match 100.0%: Score 86; DB 24: Length 319;  
 Best Local Similarity 100.0%: Pred. No. 3.80e-03;  
 Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 91 spalnkmfcq1 101  
 1 SPALNKMFCOL 11

RESULT 7  
 ID W28496 standard; Protein: 319 AA.  
 AC W28496:  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_difference 145  
 FT /note= "Arg residue at position 182 of wild-type  
 FT p53 has been mutated to His"  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1995: F01111.  
 PR 19-JUL-1995: FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.

PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 38; Page 7; 133pp: French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360-325H and comprising  
 CC the 325-360 domain, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 360-325).  
 SQ Sequence 319 AA:

Query Match 100.0%: Score 86; DB 24: Length 319;  
 Best Local Similarity 100.0%: Pred. No. 3.80e-03;  
 Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|                       |  |                            |                                 |
|-----------------------|--|----------------------------|---------------------------------|
| Db                    | 91   | spalnkmcq1 101             |                                 |
| OY                    | 1  | SPALNKMCQ1 11              |                                 |
| RESULT                | 8  |                            |                                 |
| ID                    | W28497   | standard; Protein: 335 AA. |                                 |
| AC                    | W28497   |                            |                                 |
| DT                    | 25-NOV-1997  | (first entry)              |                                 |
| DE                    | Human p53 protein variant 360h-325 encoded by PEC179.                          |                            |                                 |
| KW                    | leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;            |                            |                                 |
| KW                    | substitution; replacement; transactivation; hinge region;                      |                            |                                 |
| KW                    | anti-oncogene; hyperproliferation; cancer; restenosis;                         |                            |                                 |
| KW                    | tumour suppression; apoptosis.   |                            |                                 |
| OS                    | Homio sapiens.   |                            |                                 |
| OS                    | Synthetic.   |                            |                                 |
| FH                    | Key  | Location/Qualifiers        |                                 |
| FT                    | region   | 39..53                     |                                 |
| FT                    |  | /label= hinge              |                                 |
| PN                    | MO9704092-A1.  |                            |                                 |
| PD                    | 06-FEB-1997.   |                            |                                 |
| PF                    | 17-JUL-1996; F01111.   |                            |                                 |
| PR                    | 19-JUL-1995; FR-008729.  |                            |                                 |
| PA                    | (RHON) RHONE POULENC RORER SA.   |                            |                                 |
| PI                    | Bracco L, Conseiller E; .  |                            |                                 |
| PI                    | WPI: 97-132633/12.   |                            |                                 |
| DR                    | N-PSDB: T66224.  |                            |                                 |
| PT                    | New p53 variants e.g. with oligomerisation domain replaced by                  |                            |                                 |
| PT                    | leucine zipper. useful for treating hyper-proliferative disorders,             |                            |                                 |
| PT                    | esp. cancer and restenosis   |                            |                                 |
| PS                    | Claim 39; Pages 94-95; 133pp; French.  |                            |                                 |
| CC                    | Claimed variants of protein p53 have at least part of the                      |                            |                                 |
| CC                    | oligomerisation domain deleted and replaced by a leucine zipper                |                            |                                 |
| CC                    | domain. The mutants preferably also have at least part of the p53              |                            |                                 |
| CC                    | transactivation domain (amino acids 1-74) deleted and replaced by              |                            |                                 |
| CC                    | the domain 325-360 of p53. The present sequence is that of a                   |                            |                                 |
| CC                    | specifically claimed p53 variant designated 360h-325 and comprising            |                            |                                 |
| CC                    | the 325-360 domain, separated from amino acids 75-325 of human                 |                            |                                 |
| CC                    | wild-type p53 by a synthetic hinge sequence (GlySer) <sub>3</sub> , and with a |                            |                                 |
| CC                    | leucine zipper domain at the C-terminal. The p53 variants are                  |                            |                                 |
| CC                    | more active and more stable tumour suppressors and apoptosis-inducing          |                            |                                 |
| CC                    | agents than wild-type p53 and are active where the wild-type protein           |                            |                                 |
| CC                    | is not, i.e. they are not inactivated by dominant negative or oncogenic        |                            |                                 |
| CC                    | mutants, nor by other cellular proteins (because the leucine zipper            |                            |                                 |
| CC                    | domain prevents formation of inactive mixed oligomers).                        |                            |                                 |
| CC                    | Sequence 335 AA;   |                            |                                 |
| CC                    | 350  |                            |                                 |
| Query Match           | 100.0%;  | Score 86;                  | DB 24; Length 335;              |
| Best Local Similarity | 100.0%;  | Pred. No. 3,80e-03;        |                                 |
| Matches               | 11; Conservative   | 0;                         | Mismatches 0; Indels 0; Gaps 0; |
| Db                    | 107  | spalnkmcq1 117             |                                 |
| OY                    | 1  | SPALNKMCQ1 11              |                                 |
| RESULT                | 9  |                            |                                 |
| ID                    | W28498   | standard; Protein: 335 AA. |                                 |
| AC                    | W28498   |                            |                                 |
| DT                    | 25-NOV-1997  | (first entry)              |                                 |
| DE                    | Human p53 protein variant 360h-325H.   |                            |                                 |
| KW                    | leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;            |                            |                                 |
| KW                    | substitution; replacement; transactivation; hinge region;                      |                            |                                 |
| KW                    | anti-oncogene; hyperproliferation; cancer; restenosis;                         |                            |                                 |
| KW                    | tumour suppression; apoptosis.   |                            |                                 |
| OS                    | Homio sapiens.   |                            |                                 |
| OS                    | Synthetic.   |                            |                                 |
| FH                    | Key  | Location/Qualifiers        |                                 |
| FT                    | region   | 39..53                     |                                 |
| FT                    |  | /label= hinge              |                                 |
| FT                    | misc_difference 161  |                            |                                 |
| FT                    | /note= "Arg residue at position 182 of wild-type                               |                            |                                 |

```

FT      p53 has been mutated to His"
PD      WO9704092-A1.
PF      06-FEB-1997.
PR      17-JUL-1996; F01111.
PA      (RHON ) RHONE POULENC RORER SA.
PI      Bracco L, Consellier E;
PT      WPI: 97-132633/12.
PT      New p53 variants e.g. with oligomerisation domain replaced by
PT      leucine zipper - useful for treating hyper-proliferative disorders,
PT      esp. cancer and restenosis
PS      Claim 39; Page -: 133pp; French.
CC      Claimed variants of protein p53 have at least part of the
CC      oligomerisation domain deleted and replaced by a leucine zipper
CC      domain. The mutants preferably also have at least part of the p53
CC      transactivation domain (amino acids 1-74) deleted and replaced by
CC      the domain 325-360 of p53. The present sequence is that of a
CC      specifically claimed p53 variant designated 360h-325h and comprising
CC      the 325-360 domain, separated from amino acids 75-325 of human
CC      wild-type p53 (but with Arg182 replaced by His) by a synthetic hinge
CC      sequence (Gly4Ser3), and with a leucine zipper domain at the C-terminal
CC      The p53 variants are more active and more stable tumour suppressors
CC      and apoptosis-inducing agents than wild-type p53 and are active where
CC      the wild-type protein is not, i.e. they are not inactivated by dominant
CC      negative or oncogenic mutants, nor by other cellular proteins (because
CC      the leucine zipper domain prevents formation of inactive mixed
CC      oligomers).
CC      (Note: this sequence does not appear in the specification and has
CC      been produced by modifying the given sequence of variant 360h-325).
SQ      Sequence    335 AA;

Query Match          100.0%; Score 86; DB 24; Length 335;
Best Local Similarity 100.0%; Pred. No. 3,80e-03;
Matches   11; Conservative     0; Mismatches 0; Indels   0; Gaps   0;

Db       107 spalnkmfcql 117
        |||||
Qy       1 SPALNKMFCQL 11

RESULT      10
ID      W28494 standard; Protein; 353 AA.
AC      W28494.
DT      25-NOV-1997 (first entry)
DE      Human p53 protein variant 393-325H.
KW      Leucine zipper domain; LZD; oligomerisation domain; mutant; mutlein;
KW      substitution; replacement; transactivation; viral protein VP16; HSV;
KW      anti-oncogene; hyperproliferation; cancer; restenosis;
KW      tumour suppression; apoptosis.
OS      Homo sapiens.
OS      Synthetic.
FH      Key
FT      misc_difference 179
FT      Location/Qualifiers
FT      /note="Arg residue at position 182 of wild-type
FT      p53 has been mutated to His"
PN      WO9704092-A1.
PD      06-FEB-1997.
PF      17-JUL-1996; F01111.
PR      19-JUL-1995; FR-008729.
PA      (RHON ) RHONE POULENC RORER SA.
PI      Bracco L, Consellier E;
PT      WPI: 97-132633/12.
PT      New p53 variants e.g. with oligomerisation domain replaced by
PT      leucine zipper - useful for treating hyper-proliferative disorders,
PT      esp. cancer and restenosis
PS      Claim 37; Page -, 133pp; French.
CC      Claimed variants of protein p53 have at least part of the
CC      oligomerisation domain deleted and replaced by a leucine zipper
CC      domain. The mutants preferably also have at least part of the p53
CC      transactivation domain (amino acids 1-74) deleted and replaced by
CC      the domain 325-393 of p53. The present sequence is that of a
CC      specifically claimed p53 variant designated 393-325h and comprising
CC      the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with

```

CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 393-325).  
 SQ Sequence 353 AA;

Query Match 100.0%; Score 86; DB 24; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 3,80e-03;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dh 125 spalnkmfcql 135  
 |||||

Qy 1 SPALNKMFCOL 11

RESULT 11

ID W28493 standard; Protein; 353 AA.

AC W28493;

DT 25-NOV-1997 (first entry)

DE Human p53 protein variant 393-325 encoded by p53177.

KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;

KW substitution; replacement; transactivation; viral protein V16; HSV;

KW anti-oncogene; hyperproliferation; cancer; restenosis;

KW tumour suppression; apoptosis.

OS Homo sapiens.

OS Synthetic.

PN MO9704092-A1.

PD 06-FEB-1997.

PF 17-JUL-1996; F01111.

PR 19-JUL-1995; FR-008729.

PA (RHON ) RHONE POULENC RORER SA.

PI Biacco L, Conseiller E;

DR WPI: 97-132633/12.

N-PSDB; 786222.

PT New p53 variants e.g. with oligomerisation domain replaced by

PT leucine zipper - useful for treating hyper-proliferative disorders,

PT esp. cancer and restenosis

PS Claim 37: Pages 90-92: 133pp: French.

CC Claimed variants of protein p53 have at least part of the

CC oligomerisation domain deleted and replaced by a leucine zipper

CC domain. The mutants preferably also have at least part of the p53

CC transactivation domain (amino acids 1-74) deleted and replaced by

CC the domain 325-393 of p53. The present sequence is that of

CC a specifically claimed p53 variant designated 393-325 and comprising

CC the 325-393 domain, amino acids 75-325 of human wild-type p53 and a

CC leucine zipper domain at the C-terminal. The p53 variants are

CC more active and more stable tumour suppressors and apoptosis-inducing

CC agents than wild-type p53 and are active where the wild-type protein

CC is not, i.e. they are not inactivated by dominant negative or oncogenic

CC mutants, nor by other cellular proteins (because the leucine zipper

CC domain prevents formation of inactive mixed oligomers).

CC Sequence 353 AA;

Query Match 100.0%; Score 86; DB 24; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 3,80e-03;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dh 125 spalnkmfcql 135  
 |||||

Qy 1 SPALNKMFCOL 11

RESULT 12

ID W13971 standard; Protein; 363 AA.

AC W13971;

DT 25-JUN-1997 (first entry)

DE Modified p53 variant p53R284del364-393.

KW p53; tumour suppressor; cancer; therapy; cell proliferation.

KW apoptosis; protein engineering; DNA binding.

OS Synthetic.

PN MO9710843-A1.

PD 27-MAR-1997.

PF 20-SEP-1996; U15188.

PR 22-SEP-1995; US-004802.

PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.

PI Halazonetis TD;

DR WPI: 97-202618/18.

PT R284K modified p53 protein having DNA binding ability - useful in

PT treatment of cancer

PS Example 1: 51-52; 82pp: English.

CC Modified p53 variant p53R284del364-393 (W13971) has a Thr284 to Arg

CC substn. (see also W13949) and a deletion of the C-terminal 30

CC amino acids. The T284R substitution, introduced by site-directed

CC mutagenesis of p53 DNA, provides a novel p53-DNA contact between a

CC phosphate of the DNA backbone and p53. The C-terminal deletion

CC permits in vitro DNA binding. The variant provides the means for

CC pharmacological rescue of p53 function in cancer patients. Other

CC modified p53 constructs (W13949-50, W13953-54, W13968-77) have also

CC been produced. Nucleic acids coding for modified p53 can be used

CC for cancer gene therapy.

SQ Sequence 363 AA;

Query Match 100.0%; Score 86; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 3,80e-03;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dh 127 spalnkmfcql 137  
 |||||

Qy 1 SPALNKMFCOL 11

RESULT 13

ID W13974 standard; Protein; 363 AA.

AC W13974;

DT 25-JUN-1997 (first entry)

DE Modified p53 variant p53R273del364-393.

KW p53; tumour suppressor; cancer; therapy; cell proliferation;

KW apoptosis; protein engineering; DNA binding.

OS Synthetic.

PN MO9710843-A1.

PD 27-MAR-1997.

PF 20-SEP-1996; U15188.

PR 22-SEP-1995; US-004802.

PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.

PI Halazonetis TD;

DR WPI: 97-202618/18.

PT R284K modified p53 protein having DNA binding ability - useful in

PT treatment of cancer

PS Example 1: 56-57; 82pp: English.

CC Modified p53 variant p53R273del364-393 (W13974) has the tumour-

CC derived histidine 273 mutation (see also W13952) and a deletion

CC of the C-terminal 30 amino acids of wild-type p53 (see also

CC W13948). His273 is a Class I p53 tumour mutation that affects DNA

CC binding. The C-terminal deletion, introduced by site-directed

CC mutagenesis of p53 DNA, activates the DNA binding of the p53

CC tumour mutant. This provides the means for pharmacological rescue

CC of p53 function in cancer patients. Other modified p53 constructs

CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic

CC acids coding for modified p53 can be used for cancer gene therapy.

SQ Sequence 363 AA;

Query Match 100.0%; Score 86; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 3,80e-03;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dh 127 spalnkmfcql 137  
 |||||

Qy 1 SPALNKMFCOL 11

RESULT 14  
 ID W13973 standard; Protein; 363 AA.  
 AC W13973;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53Q248R284del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 OS apoptosis; protein engineering; DNA binding.  
 PN W09710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (W13973) W13973 INST ANATOMY & BIOLOGY.  
 PI Halazoneis TD;  
 DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 treatment of cancer.  
 PS Example 1; 54-56; 82pp; English.  
 CC Modified p53 variant p53Q248R284del364-393 (W13973) has the tumour-  
 derived Gln248 mutation (see also W13951) a Thr284 to Arg substn.  
 CC (see also W13949) and a deletion of the 30 C-terminal amino acids  
 CC of wild-type p53 (W13948). Gln248 is a Class I p53 tumour mutation  
 CC that affects DNA binding. The T284R substitution, introduced by  
 CC site-directed mutagenesis of p53 DNA, provides a novel p53-DNA  
 CC contact between a phosphate of the DNA backbone and p53, and  
 CC restores DNA binding. The C-terminal deletion permits in vitro  
 CC DNA binding. The construct provides the means for pharmacological  
 CC rescue of p53 function in cancer patients. Other modified p53  
 CC constructs (W13949-50, W13953-54, W13968-77) have also been  
 CC produced. Nucleic acids coding for modified p53 can be used for  
 CC cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 86; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 3.80e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 spalnkmfcql 137  
 QY 1 SPALNKMFCQL 11

RESULT 15  
 ID W13972 standard; Protein; 363 AA.  
 AC W13972;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53Q248del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 OS apoptosis; protein engineering; DNA binding.  
 PN W09710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (W13972) W13972 INST ANATOMY & BIOLOGY.  
 PI Halazoneis TD;  
 DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 treatment of cancer.  
 PS Example 1; 53-54; 82pp; English.  
 CC Modified p53 variant p53Q248del364-393 (W13972) has the tumour-  
 derived glutamine 248 mutation (see also W13951) and a deletion  
 CC of the C-terminal 30 amino acids of wild-type p53 (see also  
 CC W13948). Gln248 is a Class I p53 tumour mutation that affects DNA  
 CC binding. The C-terminal deletion, introduced by site-directed  
 CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
 CC tumour mutant. This provides the means for pharmacological rescue  
 CC of p53 function in cancer patients. Other modified p53 constructs  
 CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
 CC acids coding for modified p53 can be used for cancer gene therapy.

SQ Sequence 363 AA;

Query Match 100.0%; Score 86; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 3.80e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 spalnkmfcql 137  
 QY 1 SPALNKMFCQL 11

Search completed: Fri Sep 11 13:32:27 1998  
 Job time : 13 secs.



(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch.p protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:32:45 1998; MasPar time 3.73 Seconds

Tabular output not generated. 107.866 Million cell updates/sec

Title: >US-08-452-843-15  
Description: (1-11) from US08452843.pep  
Perfect Score: 86  
Sequence: 1 SPALNKMFCOL 11

Scoring table: PAM 150  
Gap 15

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p156  
1:plrf 2:plrf 3:plrf 4:plrf 5:nrl3d

Statistics: Mean 24.039; Variance 33.746; scale 0.712

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description            | Pred. No. |
|------------|-------|-------------|--------|----|--------|------------------------|-----------|
| 1          | 86    | 100.0       | 191    | 5  | 1YCSA  | p53 residues 97-287,   | 8.90e-07  |
| 2          | 86    | 100.0       | 194    | 5  | 1TUPB  | tumor suppressor p53,  | 8.90e-07  |
| 3          | 86    | 100.0       | 194    | 5  | 1TSRB  | p53 tumor suppressor,  | 8.90e-07  |
| 4          | 86    | 100.0       | 195    | 5  | 1TUPC  | tumor suppressor p53,  | 8.90e-07  |
| 5          | 86    | 100.0       | 195    | 5  | 1TSRC  | p53 tumor suppressor,  | 8.90e-07  |
| 6          | 86    | 100.0       | 196    | 5  | 1TUPA  | tumor suppressor p53,  | 8.90e-07  |
| 7          | 86    | 100.0       | 196    | 5  | 1TSRA  | p53 tumor suppressor,  | 8.90e-07  |
| 8          | 86    | 100.0       | 196    | 5  | 1TSRB  | cellular tumor antigen | 8.90e-07  |
| 9          | 82    | 95.3        | 333    | 2  | S06594 | cellular tumor antigen | 7.25e-06  |
| 10         | 77    | 89.5        | 333    | 2  | S51648 | cellular tumor antigen | 9.48e-05  |
| 11         | 77    | 89.5        | 333    | 2  | JC6176 | tumor suppressor prot  | 9.48e-05  |
| 12         | 77    | 89.5        | 336    | 2  | JH0631 | cellular tumor antigen | 1.57e-04  |
| 13         | 76    | 88.4        | 363    | 2  | A29376 | cellular tumor antigen | 2.61e-04  |
| 14         | 75    | 87.2        | 336    | 2  | JH0631 | cellular tumor antigen | 2.61e-04  |
| 15         | 75    | 87.2        | 336    | 2  | JH0631 | tumor suppressor p53   | 1.90e-03  |
| 16         | 71    | 82.6        | 391    | 2  | S02192 | cellular tumor antigen | 3.43e-02  |
| 17         | 65    | 75.6        | 367    | 2  | S02192 | cellular tumor antigen | 8.75e-02  |
| 18         | 63    | 73.3        | 381    | 2  | S38824 | cellular tumor antigen | 2.05e+00  |
| 19         | 56    | 65.1        | 1839   | 1  | RRWPEM | RNA-directed RNA poly  | 3.15e+00  |
| 20         | 55    | 64.0        | 1839   | 1  | RRWPEM | probable membrane pro  | 4.84e+00  |
| 21         | 54    | 62.8        | 162    | 2  | S52608 | GTP cyclohydrolase II  | 4.84e+00  |
| 22         | 54    | 62.8        | 344    | 2  | D64620 | hemoglobin alpha chain | 1.12e+01  |
| 23         | 52    | 60.5        | 141    | 1  | HAAG   |                        |           |

|    |    |      |      |   |        |                        |          |
|----|----|------|------|---|--------|------------------------|----------|
| 24 | 52 | 60.5 | 141  | 1 | HAAR   | hemoglobin alpha chain | 1.12e+01 |
| 25 | 52 | 60.5 | 141  | 1 | HACO   | hemoglobin alpha chain | 1.12e+01 |
| 26 | 52 | 60.5 | 435  | 2 | H64782 | hypothetical protein   | 1.12e+01 |
| 27 | 51 | 59.3 | 390  | 2 | C64996 | hypothetical protein   | 1.69e+01 |
| 28 | 51 | 59.3 | 470  | 2 | S54069 | hypothetical protein   | 1.69e+01 |
| 29 | 51 | 59.3 | 475  | 1 | W2BEM4 | gene 17 protein - sal  | 1.69e+01 |
| 30 | 51 | 59.3 | 671  | 1 | VCNVC  | env polyprotein - fel  | 1.69e+01 |
| 31 | 50 | 58.1 | 106  | 2 | S20553 | cobyrinic acid a,c-di  | 2.55e+01 |
| 32 | 50 | 58.1 | 111  | 1 | A29654 | proteinase inhibitor   | 2.55e+01 |
| 33 | 50 | 58.1 | 116  | 2 | F69370 | conserved hypothetical | 2.55e+01 |
| 34 | 50 | 58.1 | 259  | 1 | WMBES2 | 28k protein - equine   | 2.55e+01 |
| 35 | 50 | 58.1 | 353  | 2 | C44221 | orf3 protein - Autogr  | 2.55e+01 |
| 36 | 50 | 58.1 | 353  | 1 | WMNP49 | 40.9k protein - Autog  | 2.55e+01 |
| 37 | 50 | 58.1 | 631  | 1 | VCNVC  | cell fusion glycoprot  | 2.55e+01 |
| 38 | 50 | 58.1 | 631  | 1 | A48346 | cell fusion glycoprot  | 2.55e+01 |
| 39 | 50 | 58.1 | 855  | 2 | S47533 | glucose-6-phosphate 1  | 2.55e+01 |
| 40 | 50 | 58.1 | 910  | 2 | S40259 | glucose-6-phosphate 1  | 2.55e+01 |
| 41 | 50 | 58.1 | 1874 | 1 | J00533 | RNA-directed RNA poly  | 2.55e+01 |
| 42 | 49 | 57.0 | 141  | 1 | HANER  | hemoglobin alpha chain | 3.80e+01 |
| 43 | 49 | 57.0 | 378  | 2 | S33994 | finger protein ZNF118  | 3.80e+01 |
| 44 | 49 | 57.0 | 587  | 2 | E65171 | hypothetical 64.0 kD   | 3.80e+01 |
| 45 | 49 | 57.0 | 878  | 2 | A55201 | meiosis-specific prot  | 3.80e+01 |

## ALIGNMENTS

| RESULT                | 1   |                |
|-----------------------|---|----------------|
| ENTRY                 | 1YCSA   | #type complete |
| ENTRY                 | p53 residues 97-287, chain A - human                          |                |
| PDB-TITLE             | p53-53bp2 complex   |                |
| ORGANISM              | #formal name Homo sapiens #common name man                    |                |
| NOTE                  | expressed in Escherichia coli, strain B121 (d3)               |                |
| REFERENCE             | A68208  |                |
| #authors              | Gorina, S.; Pavletich, N.P.                                   |                |
| #submission           | submitted to the Brookhaven Protein Data Bank, September 1996 |                |
| #cross-references     | PDB:1YCS  |                |
| #authors              | TN001216  |                |
| #journal              | Gorina, S.; Pavletich, N.P.                                   |                |
| #title                | Structure (1996) 274:1001                                     |                |
|                       | Structure of the p53 tumor suppressor bound to the ankyrin    |                |
|                       | and sh3 domains of 53bp2.                                     |                |
| REFERENCE             | TN001217  |                |
| #authors              | Naumovski, L.; Cleary, M.L.                                   |                |
| #journal              | Mol. Cell. Biol. (1996) 16:3884                               |                |
| #title                | The p53-binding protein 53bp2 also interacts with bcl2 and    |                |
|                       | impedes cell cycle progression at g2m.                        |                |
| REFERENCE             | I38604  |                |
| #authors              | Iwabuchi, K.; Bartel, P.L.; Li, B.; Marracino, R.; Fields, S. |                |
| #journal              | Proc. Natl. Acad. Sci. U.S.A. (1994) 91:6098-6102             |                |
| #title                | Two cellular proteins that bind to wild-type but not mutant   |                |
|                       | p53.  |                |
| #cross-references     | MUID:94286584   |                |
| COMMENT               | Resolution: 2.2 angstroms                                     |                |
| COMMENT               | Determination: X-ray diffraction                              |                |
| COMMENT               | R-value: 0.205  |                |
| KEYWORDS              | ankyrin repeats; anti-oncogene; complex; disease mutation     |                |
|                       | polymorphism; multigene family; nuclear protein; p53;         |                |
|                       | phosphorylation; sh3; tumor suppressor                        |                |
| FEATURE               | 9-11  |                |
| 81-85                 | #region helix (right hand 3-10) \                             |                |
| 182-190               | #region helix (right hand alpha) \                            |                |
| 14-16,45-50,          | #region helix (right hand alpha) \                            |                |
| 134-140,99-101        | #region beta sheet \  |                |
| 28-31,36-39,          | #region beta sheet \  |                |
| 168-178,155-162,      | #region beta sheet \  |                |
| 60-67,118-123,        | #region beta sheet \  |                |
| 108-111               | #region beta sheet \  |                |
| SUMMARY               | #length 191 #molecular weight 21515 #checksum 8219            |                |
| Query Match           | 100.0%; Score 86; DB 5; Length 191;                           |                |
| Best Local Similarity | 100.0%; Pred. No. 8.90e-07;                                   |                |

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db: 31 SPALNKMFCOL 41  
QY 1 SPALNKMFCOL 11

## RESULT 2

ENTRY 1TUPB #type complete  
TITLE tumor suppressor p53, chain B - human  
PDB\_TITLE tumor suppressor p53 complexed with DNA  
ORGANISM #formal\_name Homo sapiens #common\_name man  
#note expressed in Escherichia coli

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#submission submitted to the Brookhaven Protein Data Bank, July 1995  
#cross-references PDB:1TUPB  
A43072

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Pavletich, N.P.; Chambers, K.A.; Pabo, C.O.  
#journal Genes Dev. (1993) 7:2556-2564  
#title The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots.

REFERENCE  
#authors Vogelstein, B.; Kinzler, K.W.  
#journal Cell (1992) 70:523  
#note TN031795

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

KEYWORDS anti-oncogene; complex; DNA; DNA-binding protein; tumor suppressor

FEATURE  
72-75 #region helix (right hand 3-10)\  
82-86 #region helix (right hand alpha)\  
183-192 #region helix (right hand alpha)\  
15-17,46-51, #region beta sheet\  
135-141,100-103  
29-33,37-40,  
169-179,156-163,  
61-68,119-124,  
109-112

SUMMARY  
#length 194 #molecular-weight 21830 #checksum 2852

Query Match 100.0%; Score 86; DB 5; Length 194;  
Best Local Similarity 100.0%; Pred. No. 8.90e-07;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db: 32 SPALNKMFCOL 42  
QY 1 SPALNKMFCOL 11

ENTRY 1TUPC #type complete  
TITLE tumor suppressor p53, chain C - human  
PDB\_TITLE tumor suppressor p53 complexed with DNA  
ORGANISM #formal\_name Homo sapiens #common\_name man  
#note expressed in Escherichia coli

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#submission submitted to the Brookhaven Protein Data Bank, July 1995  
#cross-references PDB:1TUPC  
A43072

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Pavletich, N.P.; Chambers, K.A.; Pabo, C.O.  
#journal Genes Dev. (1993) 7:2556-2564  
#title The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots.

REFERENCE  
#authors Vogelstein, B.; Kinzler, K.W.  
#journal Cell (1992) 70:523  
#note TN031798

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.

REFERENCE  
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.  
#journal Science (1994) 265:346-355  
#title Crystal structure of a p53 tumor suppressor--DNA complex: understanding tumorigenic mutations.



```

RESULT 5
ENTRY 1TSRC #type complete
PDB_TITLE p53 tumor suppressor, chain C - human
ORGANISM #p53 core domain in complex with DNA
#formal_name Homo sapiens #common_name man
#note expressed in Escherichia coli
REFERENCE A66760
#authors Cho, Y.; Gorina, S.; Jeffrey, P.; Pavletich, N.
#submission submitted to the Brookhaven Protein Data Bank, July 1995
#cross-references PDB:1TSR
REFERENCE A43072
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.
#journal Science (1994) 265:346-355
#title Crystal structure of a p53 tumor suppressor--DNA complex:
understanding tumorigenic mutations.
COMMENT Resolution: 2.2 angstroms
Determination: X-ray diffraction
KEYWORDS anti-oncogene; complex; DNA; DNA-binding protein; tumor
suppressor
FEATURE
11-13 #region helix (right hand 3-10)\
72-75 #region helix (right hand alpha)\
83-87 #region helix (right hand alpha)\
184-191 #region helix (right hand alpha)\
16-19,47-52, #region beta sheet\
136-142,101-103
170-181,157-164,
62-69,120-125,
110-113
SUMMARY #length 195 #molecular-weight 21917 #checksum 4657
Query Match 100.0%; Score 86; DB 5; Length 195;
Best Local Similarity 100.0%; Pred. No. 8,90e-07;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 33 SPALNMFOL 43
QY 1 SPALNMFOL 11

RESULT 6
ENTRY 1TUPA #type complete
PDB_TITLE tumor suppressor p53, chain A - human
ORGANISM #tumor suppressor p53 complexed with DNA
#formal_name Homo sapiens #common_name man
#note expressed in Escherichia coli
REFERENCE A66776
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.
#submission submitted to the Brookhaven Protein Data Bank, July 1995
#cross-references PDB:1TUP
REFERENCE A43072
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.
#journal Science (1994) 265:346-355
#title Crystal structure of a p53 tumor suppressor--DNA complex:
understanding tumorigenic mutations.
REFERENCE A49450
#authors Pavletich, N.P.; Chambers, K.A.; Pabo, C.O.
#journal Genes Dev. (1993) 7:2556-2564
#title The DNA-binding domain of p53 contains the four conserved
regions and the major mutation hot spots.
TNO31792
Vogelstein, B.; Kinzler, K.W.
Cell (1992) 70:523
COMMENT Resolution: 2.2 angstroms
Determination: X-ray diffraction
KEYWORDS R-value: 0.202
antigen p53; complex; DNA; tumor suppressor
FEATURE
73-75 #region helix (right hand 3-10)\
84-87 #region helix (right hand alpha)\
185-194 #region helix (right hand alpha)\

```

```

17-19,48-53, #region beta sheet\
137-143,102-105
31-34,39-42,
171-181,158-165,
63-70,121-126,
111-114
SUMMARY #length 196 #molecular-weight 22004 #checksum 7058
Query Match 100.0%; Score 86; DB 5; Length 196;
Best Local Similarity 100.0%; Pred. No. 8,90e-07;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 34 SPALNMFOL 44
QY 1 SPALNMFOL 11

RESULT 7
ENTRY 1TSRA #type complete
PDB_TITLE p53 tumor suppressor, chain A - human
ORGANISM #p53 core domain in complex with DNA
#formal_name Homo sapiens #common_name man
#note expressed in Escherichia coli
REFERENCE A66760
#authors Cho, Y.; Gorina, S.; Jeffrey, P.; Pavletich, N.
#submission submitted to the Brookhaven Protein Data Bank, July 1995
#cross-references PDB:1TSR
REFERENCE A43072
#authors Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P.
#journal Science (1994) 265:346-355
#title Crystal structure of a p53 tumor suppressor--DNA complex:
understanding tumorigenic mutations.
COMMENT Resolution: 2.2 angstroms
Determination: X-ray diffraction
KEYWORDS anti-oncogene; complex; DNA; DNA-binding protein; tumor
suppressor
FEATURE
73-75 #region helix (right hand 3-10)\
84-87 #region helix (right hand alpha)\
185-194 #region helix (right hand alpha)\
17-19,48-53, #region beta sheet\
137-143,102-105
31-34,39-42,
171-181,158-165,
63-70,121-126,
111-114
SUMMARY #length 196 #molecular-weight 22004 #checksum 7058
Query Match 100.0%; Score 86; DB 5; Length 196;
Best Local Similarity 100.0%; Pred. No. 8,90e-07;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 34 SPALNMFOL 44
QY 1 SPALNMFOL 11

RESULT 8
ENTRY DNH053 #type complete
PDB_TITLE cellular tumor antigen p53 - human
ALTERNATE_NAMES cellular phosphoprotein p53; oncoprotein p53; transformation
suppressor p53; tumor suppressor p53
ORGANISM #formal_name Homo sapiens #common_name man
DATE 05-Oct-1988 #sequence_revision 18-Nov-1994 #text_change
18-Sep-1997
ACCESSIONS A25234; A43073; J00436; S40773; S42669; A22837; A55060;
A25397; B25397; S42452; S42453; I38082; I38083; I38084;
I38085; I38086; I38087; I38088; I38089; I38090; I38091;
I38092; I38093; A44905; I58354; I78850; S60153
A25224
REFERENCE A25224
#authors Lamb, P.; Crawford, L.
#journal Mol. Cell. Biol. (1986) 6:1379-1385
#title Characterization of the human p53 gene.

```

```
#cross-references MUID:87064416
#accession A25224
##molecule_type DNA
##residues 1-393 ##label LAM
##cross-references EMBL:X01405; GB:M13121; GB:N00032; NID:9189460;
PID:9386994
REFERENCE
#authors JT0436
#journal Buchman, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.;
Georgiev, G.P.
#title Gene (1988) 70:245-252
#title A variation in the structure of the protein-coding region of
the human p53 gene.
#cross-references MUID:89108008
#accession A43073
##molecule_type DNA
##residues 1-393 ##label BUC
#note this 72-Arg allele appears to be about 5 times more
frequent than the 72-Pro allele
#accession JT0436
##molecule_type DNA
##residues 1-71,'P',73-393 ##label B02
##cross-references EMBL:M22898; NID:9189474; PID:9189476
#note this 72-Pro allele was found in both normal and
malignant cell lines
REFERENCE
#accession S40773
#authors Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.
#journal Submitted to the EMBL Data Library, August 1990
#accession S40773
##molecule_type DNA
##residues 1-393 ##label CHU
##cross-references EMBL:X54156; NID:935213; PID:935214
S42669
REFERENCE
#authors Matlashevski, G.; Lamb, P.; Plm, D.; Peacock, J.; Crawford,
L.; Benichmol, S.
#journal EMBO J. (1984) 3:3257-3262
#title Isolation and characterization of a human p53 cDNA clone:
expression of the human p53 gene.
#accession S42669
##molecule_type mRNA
##residues 101-393 ##label MK1
##cross-references EMBL:X01405; NID:935215; PID:9642241
REFERENCE
#authors Zakut-Houri, R.; Bienez-Tadmor, B.; Givoli, D.; Oren, M.
#journal EMBO J. (1985) 4:1251-1255
#title Human p53 cellular tumor antigen: cDNA sequence and
expression in COS cells.
#cross-references MUID:85230577
#accession A22837
##molecule_type mRNA
##residues 1-71,'P',73-393 ##label ZAK
##cross-references EMBL:X02469; EMBL:M60950; NID:935209; PID:935210
REFERENCE
#authors Harlow, E.; Williamson, N.M.; Ralston, R.; Helfman, D.M.;
Adams, T.E.
#journal Mol. Cell. Biol. (1985) 5:1601-1610
#title Molecular cloning and in vitro expression of a cDNA clone for
human cellular tumor antigen p53.
#accession A55060
##molecule_type mRNA
##residues 1-71,'P',73-272,'H',274-393 ##label HA3
##cross-references GB:X03199; NID:9189478; PID:9189479
##experimental_source clone PR4-2, cell line A431
A93086
REFERENCE
#authors Harris, N.; Billi, E.; Shohat, O.; Prokocimer, M.; Wolf, D.;
Arai, N.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:4650-4656
#title Molecular basis for heterogeneity of the human p53 protein.
#cross-references MUID:87089826
#accession A25397
##molecule_type mRNA
##residues 1-78,'T',80-393 ##label HAR
##cross-references EMBL:M14694; NID:9339813; PID:9339814
##experimental_source clone p53-H-1, transformed hybridoma SV-80 cell
line
#accession B25397
##molecule_type mRNA
##residues 1-71,'P',73-78,'T',80-393 ##label HA2
##cross-references EMBL:M14695; NID:9339815; PID:9339816
##experimental_source clone p53-H-19, transformed hybridoma SV-80 cell
line
REFERENCE
#authors S42452
#journal Matlashevski, G.J.; Tuck, S.; Plm, D.; Lamb, P.; Schneider,
J.; Crawford, L.V.
#title Mol. Cell. Biol. (1987) 7:961-963
#title Primary structure polymorphism at amino acid residue 72 of
human p53.
#accession S42452
##molecule_type mRNA: DNA
##residues 66-71,'P',73-79 ##label MK2
##experimental_source clone lambda C113
#note 72-Cys was also found, and appears to represent a
polymorphism
#accession S42453
##molecule_type mRNA: DNA
##residues 66-79 ##label MAT
#note experimental_source clone J6K
REFERENCE
#authors Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.;
Crook, T.
#journal EMBO J. (1991) 10:2879-2887
#title p53 is frequently mutated in Burkitt's lymphoma cell lines.
#cross-references MUID:92007731
#accession I38082
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-189,'LTLSTSEKREKCVSWTETFLPIYWKCPRLRLALT',
VPSSTTTCVTPANAA' ##label F01
##cross-references EMBL:X60010; NID:9506432; PID:9506433
#note deletion of a C nucleotide causes a frameshift at
position 566
#accession I38083
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-192,'R',194-393 ##label F02
##cross-references EMBL:X60011; NID:9506434; PID:9506435
#accession I38084
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-393 ##label F03
##cross-references EMBL:X60012; NID:9506436; PID:9506437
#accession I38085
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-245,'T',247-393 ##label F04
##cross-references EMBL:X60013; NID:9506438; PID:9506439
#accession I38086
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-236,'T',238-393 ##label F05
##cross-references EMBL:X60014; NID:9506440; PID:9506441
#accession I38087
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-247,'Q',249-393 ##label F06
##cross-references EMBL:X60015; NID:9506442; PID:9506443
#accession I38088
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-71,'P',73-237,'Y',239-393 ##label F07
##cross-references EMBL:X60016; NID:9506444; PID:9506445
#accession I38089
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-247,'Q',249-393 ##label F08
##cross-references EMBL:X60017; NID:9506446; PID:9506447
#accession I38090
```

```

#status      translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    1-71,'P',73-162,'H',164-393 ##label F09
#cross-references EMBL:X60018; NID:9506448; PID:9506449
#accession   138091
#status      translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    1-212,'Q',214-393 ##label F10
#cross-references EMBL:X60019; NID:9506450; PID:9506451
#accession   138092
#status      translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    1-253,'D',255-393 ##label F11
#cross-references EMBL:X60020; NID:9506452; PID:9506453
#note        all sequences submitted to the EMBL/GenBank/DBJ
              databases June 1991

REFERENCE    138093
#authors     Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.
#journal     Nucleic Acids Res. (1991) 19:6977
#title       An Alu polymorphism intragenic to the TP53 gene.
#cross-references MIM:92107726
#accession   138093
#status      translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    1-393 ##label RE2
#cross-references EMBL:X54156; NID:935213; PID:935214
#accession   A44905
#authors     Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;
              Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.;
              Terada, M.
#journal     Cancer Res. (1991) 51:5800-5805
#title       p53 gene mutations in gastric cancer metastases and in
              gastric cancer cell lines derived from metastases.
#cross-references MIM:92034678
#accession   A44905
#molecule_type DNA
#residues    246-247,'W',249-250 ##label YAW
#cross-references GB:563157; NID:9237829; PID:9237830
#note        sequence extracted from NCBI Backbone (NCBIN:63157,

Note: remainder of annotations omitted.

Query Match      100.0%; Score 86; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 8.90e-07;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      127 SPALNKMFCOL 137
      1 SPALNKMFCOL 11

RESULT 9
ENTRY   S06594 #type complete
TITLE   cellular tumor antigen p53 - green monkey
ORGANISM #formal_name Cercopithecus aethiops #common_name green
          monkey, grivet
DATE    28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
          08-Sep-1997
ACCESSIONS S06594
REFERENCE   S06594
#authors    Rigaudy, P.; Eckhart, W.
#journal     Nucleic Acids Res. (1989) 17:8375
#title       Nucleotide sequence of a cDNA encoding the monkey cellular
              phosphoprotein p53.
#cross-references MIM:90045967
#accession   S06594
#molecule_type mRNA
#residues    1-393 ##label RIG
#cross-references EMBL:X16384; NID:922795; PID:922796
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS      apoptosis; cell division control; DNA binding; homotetramer;
              nucleus; phosphoprotein; transcription regulation; tumor
              suppressor; zinc

```

```

FEATURE
116,179,238,242 #binding_site zinc (Cys, His, Cys, Cys) #status
392             predicted\
              #binding_site phosphoryl-RNA (Ser) (covalent) #status
              predicted
SUMMARY        #length 393 #molecular-weight 43696 #checksum 4263

Query Match      95.3%; Score 82; DB 2; Length 393;
Best Local Similarity 90.9%; Pred. No. 7.25e-06;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db      127 SPDLNKMFCOL 137
      1 SPALNKMFCOL 11

RESULT 10
ENTRY   S51648 #type complete
TITLE   cellular tumor antigen p53 - bovine
ALTERNATE_NAMES tumor-suppressor protein p53
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE    07-May-1995 #sequence_revision 01-Sep-1995 #text_change
          08-Sep-1997
ACCESSIONS S51648
REFERENCE   S51648
#authors    Degueldt, F.; Willems, L.; Burny, A.; Kettmann, R.
#journal     Submitted to the EMBL Data Library, September 1994
#title       Nucleotide sequence of the ovine p53 tumor-suppressor gene
              cDNA and its genomic organisation.
#accession   S51648
#status      preliminary
#molecule_type mRNA
#residues    1-386 ##label DEQ
#cross-references EMBL:X81704; NID:9602332; PID:9602333
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS      apoptosis; cell division control; DNA binding; homotetramer;
              phosphoprotein; transcription regulation; tumor suppressor;
              zinc

FEATURE
168,171,231,235 #binding_site zinc (Cys, His, Cys, Cys) #status
385             predicted\
              #binding_site phosphoryl-RNA (Ser) (covalent) #status
              predicted
SUMMARY        #length 386 #molecular-weight 43255 #checksum 7025

Query Match      89.5%; Score 77; DB 2; Length 386;
Best Local Similarity 81.8%; Pred. No. 9.48e-05;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db      119 SPSLNKLFQOL 129
      1 SPALNKMFCOL 11

RESULT 11
ENTRY   JC6176 #type complete
TITLE   tumor suppressor protein p53 - Chinese hamster
ORGANISM #formal_name Crictulius griseus #common_name Chinese hamster
DATE    11-Apr-1997 #sequence_revision 09-May-1997 #text_change
          08-Sep-1997
ACCESSIONS JC6176
REFERENCE   JC6176
#authors    Lee, H.; Larner, J.M.; Hamlin, J.L.
#journal     Gene (1997) 184:177-183
#title       Cloning and characterization of Chinese hamster p53 cDNA.
              liver
#accession   JC6176
#molecule_type mRNA
#residues    1-393 ##label LEE
#cross-references GB:U50395; NID:91842229; PID:91842230
COMMENT      This protein is a multimer. It plays the central role in a complex
              DNA damage-sensing network. It binds to replication factor and
              TATA-binding protein, and affects DNA replication, transcription.

```

## and recombination by protein/protein interactions.

GENETICS and recombination by protein/protein interactions.  
 #gene p53  
 CLASSIFICATION #superfamily cellular tumor antigen p53  
 KEYWORDS liver; tumor  
 SUMMARY #length 393 #molecular-weight 43362 #checksum 4043

Query Match 89.5%; Score 77; DB 2; Length 393;  
 Best Local Similarity 81.8%; Pred. No. 9,48e-05;  
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 127:SPSLNKLFCOL 137  
 11:111:1111  
 QY 1 SPALNKLFCOL 11

RESULT 12  
 ENTRY JH0633 #type complete  
 TITLE cellular tumor antigen p53 - golden hamster  
 ALTERNATE\_NAMES tumor-suppressor protein p53  
 ORGANISM #formal\_name Mesocricetus auratus #common\_name golden hamster  
 DATE 17-Aug-1992 #sequence\_revision 17-Aug-1992 #text\_change 08-Sep-1997

ACCESSIONS JH0633  
 REFERENCE JH0633  
 #authors Legros, Y.; McIntyre, P.; Soussi, T.  
 #journal Gene (1992) 112:247-250  
 #title The cDNA cloning and immunological characterization of hamster p53.

#cross-references MUID:92210007  
 #accession JH0633  
 #molecule\_type mRNA  
 #residues 1-396 #label LEG  
 ##cross-references GB:M75144; NID:q191414; PID:q191415  
 ##experimental\_source Kidney, strain MPl

GENETICS #gene p53  
 CLASSIFICATION #superfamily cellular tumor antigen p53  
 KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; nucleus; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 179,182,241,245 #binding\_site zinc (Cys, His, Cys, Cys) #status predicted  
 395 #binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted  
 SUMMARY #length 396 #molecular-weight 43631 #checksum 6617

Query Match 89.5%; Score 77; DB 2; Length 396;  
 Best Local Similarity 81.8%; Pred. No. 9,48e-05;  
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 130:SPSLNKLFCOL 140  
 11:111:1111  
 QY 1 SPALNKLFCOL 11

RESULT 13  
 ENTRY DNMS53 #type complete  
 TITLE cellular tumor antigen p53 - mouse  
 ALTERNATE\_NAMES oncoprotein p53  
 ORGANISM #formal\_name Mus musculus #common\_name house mouse  
 DATE 28-Aug-1985 #sequence\_revision 04-Oct-1996 #text\_change 05-Sep-1997

ACCESSIONS A22739; S06336; A02684; S38822; S38823; I48703  
 REFERENCE A22739  
 #authors Bianz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.  
 #journal EMBO J. (1984) 3:2179-2183  
 #cross-references MUID:85027173  
 #accession A22739  
 #molecule\_type DNA  
 #residues 1-134, 'V', 136-390 #label BIE

REFERENCE S06336

#authors Chumakov, P.M.  
 #journal Bloorg. Khim. (1987) 13:1691-1694  
 #title Primary structure of DNA complementary to murine oncoprotein p53 mRNA.  
 #cross-references MUID:88221682  
 #accession S06336  
 #status not compared with conceptual translation

REFERENCE #molecule\_type mRNA  
 #residues 1-134, 'V', 136-390 #label CHU  
 #authors Zakut-Houri, R.; Oren, M.; Bianz, B.; Lavie, V.; Hazum, S.; Givol, D.

#journal Nature (1983) 306:594-597  
 #title A single gene and a pseudogene for the cellular tumour antigen p53.  
 #cross-references MUID:84068204  
 #accession A02684  
 #molecule\_type mRNA

REFERENCE #residues 1-159, 'H', 161-167, 'G', 169-233, 'I', 235-390 #label ZAK  
 #authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.; Shohat, O.; Rotter, V.  
 #journal Mol. Cell. Biol. (1986) 6:3232-3239  
 #title Immunologically distinct p53 molecules generated by alternative splicing.

#accession S38822  
 #status preliminary  
 #molecule\_type mRNA  
 #residues 1-390 #label ARA  
 ##cross-references EMBL:M13872; NID:9200198; PID:9200199  
 #accession S38823

REFERENCE #status preliminary  
 #molecule\_type mRNA  
 #residues 1-167, 'G', 169-233, 'I', 235-390 #label AR2  
 ##cross-references EMBL:M13873

REFERENCE #authors Jenkins, J.R.; Ruddle, K.; Redmond, S.; Wade-Evans, A.  
 #journal Nucleic Acids Res. (1984) 12:5609-5626  
 #title Cloning and expression analysis of full length mouse cDNA sequences encoding the transformation associated protein p53.

#cross-references MUID:84272240  
 #accession I48703  
 #status preliminary; translated from GB/EMBL/DBD

COMMENT #molecule\_type mRNA  
 #residues 1-47, 'R', 49-78, 'QW', 82-390 #label RBS  
 ##cross-references EMBL:X00741; NID:953570; PID:953571  
 COMMENT This DNA-binding protein plays an essential role in the regulation of cell division, as it is required for the transition from phase G0 to G1 of the cell cycle.

COMMENT The tetramer association region may exhibit a beta-turn, beta-sheet, beta-turn, alpha-helix motif.  
 CLASSIFICATION #superfamily cellular tumor antigen p53  
 KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 1-44  
 16-26 #domain transcription activation #status predicted  
 99-289 #label TRA  
 108-111 #region conserved region I\  
 114-139 #region L1 loop\  
 160-192 #region conserved region II\  
 168-178 #region L2 loop\  
 231-252 #region conserved region III\  
 233-248 #region conserved region IV\  
 267-283 #region L3 loop\  
 313-319 #region conserved region V\  
 313-319 #region nuclear location signal\  
 319-357 #region tetramer association\  
 7,9,12,18,23,37 #binding\_site phosphate (Ser) (covalent) #status predicted\  
 173,176,235,239 #binding\_site zinc (Cys, His, Cys, Cys) #status predicted

Sun Sep 13 10:55:34 1998

```

312      predicted\
      #binding_site phosphate (Ser) (covalent) (by cdc2
      kinase) #status predicted\
389      #binding_site phosphoryl-RNA (Ser) (covalent) #status
SUMMARY      #length 300 #molecular-weight 43458 #checksum 1260

Query Match      88.4%; Score 76; DB 1; Length 390;
Best Local Similarity 81.8%; Pred. No. 1,57e-04;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 124:SPLNKLFQOL 134
OY 1 SPALNMFQOL 11

RESULT 14
ENTRY      A29376      #type complete
TITLE      cellular tumor antigen p53 - African clawed frog
ORGANISM   #formal_name Xenopus laevis #common_name African clawed frog
DATE       31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change
08-Sep-1997
ACCESSIONS A29376; S61531; S72313; I51639
REFERENCE   A29376
#authors    Soussi, T.; de Fromental, C.C.; Mechali, M.; May, P.; Kress,
M.
#journal    Oncogene (1987) 1:71-78
#title      Cloning and characterization of a cDNA from Xenopus laevis
coding for a protein homologous to human and murine p53.
#cross-references M0ID:88143684
#accession  A29376
#molecule_type mRNA
#residues   1-363 #label SOU
REFERENCE   I51639
#cross-references EMBL:X05191; NID:964961; PID:964962
#authors    Hoever, M.; Clement, J.H.; Wedlich, D.; Montenarh, M.;
Knoechel, W.
#journal    Oncogene (1994) 9:109-120
#title      Overexpression of wild-type p53 interferes with normal
development in Xenopus laevis embryos.
#cross-references M0ID:94134403
#accession  S61531
#molecule_type mRNA
#residues   1-293,295-363 #label HOE
REFERENCE   S72313
#cross-references EMBL:X77546; NID:9468513; PID:9468514
#authors    Hoever, M.; Clement, J.; Wedlich, D.; Montenarh, M.; Knoechel,
M.
#submission submitted to the EMBL Data Library, March 1994
#accession  S72313
#molecule_type mRNA
#residues   1-51,'S',53-70,72-293,295-363 #label HOW
#cross-references EMBL:X77546; NID:9468513; PID:9468514
GENETICS     p53
#gene        #superfamily cellular tumor antigen p53
CLASSIFICATION apoptosis; cell division control; DNA binding; homotetramer;
KEYWORDS      nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc

FEATURE
150,153,213,217      #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
362      #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted
SUMMARY      #length 363 #molecular-weight 40692 #checksum 6648

Query Match      87.2%; Score 75; DB 2; Length 363;
Best Local Similarity 81.8%; Pred. No. 2,61e-04;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 101:SPLNKLFQOL 111
OY 1 SPALNMFQOL 11

```

```

RESULT 15
ENTRY      JH0631      #type complete
TITLE      cellular tumor antigen p53 - rainbow trout
ORGANISM   #formal_name Oncorhynchus mykiss #common_name rainbow trout
DATE       17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
08-Sep-1997
ACCESSIONS JH0631
REFERENCE   JH0631
#authors    de Fromental, C.C.; Pakdel, F.; Chapus, A.; Baney, C.; May,
P.; Soussi, T.
#journal    Gene (1992) 112:241-245
#title      Rainbow trout p53: cDNA cloning and biochemical
characterization.
#cross-references M0ID:92210006
#accession  JH0631
#molecule_type mRNA
#residues   1-396 #label DEF
#cross-references GB:M75145; NID:9213828; PID:9213829
#experimental_source liver
COMMENT      This protein is the product of a tumor suppressor gene, p53, whose
inactivation leads to cell transformation or neoplasia.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS      apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc

FEATURE
164,167,227,231      #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
395      #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted
SUMMARY      #length 396 #molecular-weight 43966 #checksum 9018

Query Match      87.2%; Score 75; DB 2; Length 396;
Best Local Similarity 81.8%; Pred. No. 2,61e-04;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 115:SPDNKLFQOL 125
OY 1 SPALNMFQOL 11

```

Search completed: Fri Sep 11 13:33:00 1998  
Job time : 15 secs.



\*\*\*\*\*  
 WISE  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSrch\_p protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Sep 11 13:33:18 1998; MasPar time 2.53 Seconds  
 Tabular output not generated. 108.975 Million cell updates/sec

Title: >US-08-452-843-15  
 Description: (1-11) from US08452843.pap  
 Perfect Score: 86  
 Sequence: 1 SPALNMFQOL 11

Scoring table:  
 PAM 150  
 Gap 15

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: 1 swiss-prot35  
 1:swiss1

Statistics: Mean 25.109; Variance 29.224; scale 0.859

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID        | Description            | Pred. No. |
|------------|-------|-------------|--------|-----------|------------------------|-----------|
| 1          | 86    | 100.0       | 207    | P53_EOVAS | CELLULAR TUMOR ANTIGEN | 1.55e-08  |
| 2          | 86    | 100.0       | 393    | P53_HUMAN | CELLULAR TUMOR ANTIGEN | 1.55e-08  |
| 3          | 82    | 95.3        | 393    | P53_CERAE | CELLULAR TUMOR ANTIGEN | 1.84e-07  |
| 4          | 77    | 89.5        | 280    | P53_HORSE | CELLULAR TUMOR ANTIGEN | 3.79e-06  |
| 5          | 77    | 89.5        | 314    | P53_SPERE | CELLULAR TUMOR ANTIGEN | 3.79e-06  |
| 6          | 77    | 89.5        | 382    | P53_SHEEP | CELLULAR TUMOR ANTIGEN | 3.79e-06  |
| 7          | 77    | 89.5        | 386    | P53_BOVIN | CELLULAR TUMOR ANTIGEN | 3.79e-06  |
| 8          | 77    | 89.5        | 393    | P53_CROGR | CELLULAR TUMOR ANTIGEN | 3.79e-06  |
| 9          | 77    | 89.5        | 396    | P53_MESAU | CELLULAR TUMOR ANTIGEN | 3.79e-06  |
| 10         | 76    | 88.4        | 386    | P53_FELCA | CELLULAR TUMOR ANTIGEN | 6.89e-06  |
| 11         | 75    | 88.4        | 390    | P53_MOUSE | CELLULAR TUMOR ANTIGEN | 6.89e-06  |
| 12         | 75    | 87.2        | 363    | P53_XENLA | CELLULAR TUMOR ANTIGEN | 1.25e-05  |
| 13         | 75    | 87.2        | 373    | P53_BRARE | CELLULAR TUMOR ANTIGEN | 1.25e-05  |
| 14         | 75    | 87.2        | 396    | P53_SALIR | CELLULAR TUMOR ANTIGEN | 1.25e-05  |
| 15         | 72    | 83.7        | 276    | P53_CANFA | CELLULAR TUMOR ANTIGEN | 7.21e-05  |
| 16         | 71    | 82.6        | 391    | P53_RABIT | CELLULAR TUMOR ANTIGEN | 1.29e-04  |
| 17         | 65    | 75.6        | 391    | P53_PAT   | CELLULAR TUMOR ANTIGEN | 3.79e-03  |
| 18         | 63    | 73.3        | 351    | P53_ORILA | CELLULAR TUMOR ANTIGEN | 1.13e-02  |
| 19         | 63    | 73.3        | 367    | P53_CHICK | CELLULAR TUMOR ANTIGEN | 1.13e-02  |
| 20         | 55    | 64.0        | 366    | P53_PLAFE | CELLULAR TUMOR ANTIGEN | 7.27e-01  |
| 21         | 55    | 64.0        | 1839   | P53_PLMV  | RNA REPLICASE POLYPROT | 7.27e-01  |
| 22         | 52    | 60.5        | 141    | HBA_CRONI | HEMOGLOBIN ALPHA CHAIN | 3.14e+00  |
| 23         | 52    | 60.5        | 141    | HBA_CAIOR | HEMOGLOBIN ALPHA CHAIN | 3.14e+00  |

| Result | ID | Score | Query Match | Length | ID         | Description            | Pred. No. |
|--------|----|-------|-------------|--------|------------|------------------------|-----------|
| 24     | 52 | 60.5  | 141         | 1      | HBA_ALIMI  | HEMOGLOBIN ALPHA CHAIN | 3.14e+00  |
| 25     | 52 | 60.5  | 351         | 1      | HOMN_ALCED | HIGH-AFFINITY NICKEL T | 3.14e+00  |
| 26     | 51 | 59.3  | 475         | 1      | VP40_HSVSA | CAPSID PROTEIN P40 (CO | 5.05e+00  |
| 27     | 51 | 59.3  | 671         | 1      | ENV_FENVI  | ENV POLYPROTEIN PRECUR | 5.05e+00  |
| 28     | 50 | 58.1  | 111         | 1      | IPSG_FELCA | DOUBLE-HEADED PROTEASE | 8.06e+00  |
| 29     | 50 | 58.1  | 259         | 1      | US10_HSV4  | 28 KD PROTEIN (ORF3)   | 8.06e+00  |
| 30     | 50 | 58.1  | 347         | 1      | U133_HSV6  | G-PROTEIN COUPLED RECE | 8.06e+00  |
| 31     | 50 | 58.1  | 353         | 1      | VEA1_NPVAC | EARLY 40.9 KD PROTEIN  | 8.06e+00  |
| 32     | 50 | 58.1  | 459         | 1      | CBIA_SALTY | COBRINIC ACID A,C-DIA  | 8.06e+00  |
| 33     | 50 | 58.1  | 631         | 1      | VGLE_PHOBY | FUSION GLYCOPROTEIN PR | 8.06e+00  |
| 34     | 50 | 58.1  | 1874        | 1      | POLR_KYMWJ | RNA REPLICASE POLYPROT | 8.06e+00  |
| 35     | 49 | 57.0  | 119         | 1      | Y13K_NPVOP | HYPOTHETICAL 13.4 KD P | 1.28e+01  |
| 36     | 49 | 57.0  | 141         | 1      | HBA1_PLEMA | HEMOGLOBIN ALPHA-1 CHA | 1.28e+01  |
| 37     | 49 | 57.0  | 186         | 1      | INVB_PARBA | INVASION PROTEIN B.    | 1.28e+01  |
| 38     | 49 | 57.0  | 379         | 1      | Z11B_HUMAN | ZINC FINGER PROTEIN B. | 1.28e+01  |
| 39     | 49 | 57.0  | 566         | 1      | SYRD_PSESY | ATP-BINDING PROTEIN SY | 1.28e+01  |
| 40     | 49 | 57.0  | 587         | 1      | YIDU_ECOLI | HYPOTHETICAL 64.0 KD P | 1.28e+01  |
| 41     | 49 | 57.0  | 848         | 1      | Z33A_HUMAN | ZINC FINGER PROTEIN 33 | 1.28e+01  |
| 42     | 49 | 57.0  | 878         | 1      | MSH4_YEAST | MUTS PROTEIN HOMOLOG 4 | 1.28e+01  |
| 43     | 48 | 55.8  | 278         | 1      | OXZG_RAT   | OX-2 MEMBRANE GLYCOPRO | 2.01e+01  |
| 44     | 48 | 55.8  | 366         | 1      | YGSF_YEAST | PUTATIVE MITOCHONDRIAL | 2.01e+01  |
| 45     | 48 | 55.8  | 507         | 1      | CPT7_MOUSE | CYTOCHROME P450 XVIIA1 | 2.01e+01  |

## ALIGNMENTS

| Result | ID            | Score  | Query Match   | Length | ID           | Description | Pred. No. |
|--------|---------------|--------|---|--------|--------------|-------------|-----------|
| 1      | P53_EOVAS     | 100.0% | Score 86; DB 1; Length 207; Best Local Similarity 100.0%; Pred. No. 1.55e-08; Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | 207    | AA: DPBAE9C1 | CRC32       |           |
| 2      | SPALNMFQOL 12 | 100.0% | Score 86; DB 1; Length 207; Best Local Similarity 100.0%; Pred. No. 1.55e-08; Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | 207    | AA: DPBAE9C1 | CRC32       |           |
| 3      | SPALNMFQOL 11 | 100.0% | Score 86; DB 1; Length 207; Best Local Similarity 100.0%; Pred. No. 1.55e-08; Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | 207    | AA: DPBAE9C1 | CRC32       |           |

RESULT 2  
ID P53\_HUMAN STANDARD: PRT: 393 AA.  
AC P04637;  
DT 13-AUG-1987 (REL. 05, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
GN TP53.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC BUTHIRIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85230577.  
RA ZAKUT-HOURI R., BIENZ-TADMOR B., GIVOL D., OREN M.;  
EMBO J. 4:1251-1255(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87064416.  
RA LAMB P., CRAWFORD L.;  
MOL. CELL. BIOL. 6:1379-1385(1986).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85267676.  
RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;  
MOL. CELL. BIOL. 5:1601-1610(1985).  
RN [4]  
RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
RX MEDLINE: 87089826.  
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
ROTTER V.;  
MOL. CELL. BIOL. 6:4650-4656(1986).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 89108008.  
RA BUCHANAN V.L., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P.,  
GEOREGIEV G.P.;  
GENE 70:245-252(1988).  
RN [6]  
RP SEQUENCE OF 101-393 FROM N.A.  
RX MEDLINE: 85126934.  
RA MATLASHENSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
BENTHIMOL S.;  
EMBO J. 3:3257-3262(1984).  
RN [7]  
RP NUCLEAR LOCALIZATION SIGNAL.  
RX ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
ONCOGENE 5:423-426(1990).  
RN [8]  
RP STRUCTURE BY NMR OF 319-360.  
RX MEDLINE: 94294808.  
RA CLORE G.M., OMICHENSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
APPELLA E., GRONENBORN A.M.;  
SCIENCE 265:386-391(1994).  
RN [9]  
RP STRUCTURE BY NMR OF 325-355.  
RX MEDLINE: 95292092.  
RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
NAT. STRUCT. BIOL. 1:877-890(1994).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
RX MEDLINE: 94294806.  
RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
SCIENCE 265:346-355(1994).  
RN [11]  
RP REVIEW.  
RX MEDLINE: 94090335.  
RA HARRIS C.C.;  
SCIENCE 262:1980-1981(1993).  
RN [12]

RP REVIEW ON VARIANTS.  
RX MEDLINE: 91289156.  
RA HOULSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
SCIENCE 253:49-53(1991).  
RN [13]  
RP REVIEW ON VARIANTS.  
RX MEDLINE: 96271983.  
RA DE VRIES E.M.G., RICE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
LIAO D., SOUSSI T., KOVACH J.S., SOMMER S.S.;  
HUM. MUTAT. 7:202-213(1996).  
RN [14]  
RP VARIANT ARG-72.  
RX MEDLINE: 91153807.  
RA OLSCHWANG S., LAUBENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
HDM. GENET. 86:369-370(1991).  
RN [15]  
RP VARIANT LFS THR-133.  
RX MEDLINE: 92034774.  
RA LAM J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
CANCER RES. 51:6385-6387(1991).  
RN [16]  
RP VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
RX MEDLINE: 91057657.  
RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
KIM D.H., KASSEL J., GRAY M.A., BISCHOFF F.Z., TAINSKY M.A.,  
FRIEND S.H.;  
SCIENCE 250:1233-1238(1990).  
RN [17]  
RP VARIANT LFS ASP-245.  
RX MEDLINE: 91080929.  
RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
NATURE 348:747-749(1990).  
RN [18]  
RP VARIANT LFS LEU-272.  
RX MEDLINE: 92147883.  
RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
KNUDSEN T., MINNA J.D.;  
J. CLIN. INVEST. 89:640-647(1992).  
RN [19]  
RP VARIANTS LFS HIS-273 AND VAL-325.  
RX MEDLINE: 92228023.  
RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
GERHARDT M.C., ANDERSEN T.I., BORRESSEN A.-L., LI F.P., GARBER J.,  
STRONG L.C.;  
NEW ENGL. J. MED. 326:1309-1315(1992).  
RN [20]  
RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
RX MEDLINE: 90295284.  
RA BARTER J., IGGO R., GANNON J., LANE D.P.;  
ONCOGENE 5:893-899(1990).  
RN [21]  
RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
RX MEDLINE: 91017544.  
RA RODRIGUES N.R., ROMAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
GANNON J.V., LANE D.P.;  
PROC. NATL. ACAD. SCI. U.S.A. 87:7555-7559(1990).  
RN [22]  
RP VARIANTS COLORECTAL CANCER MUTATIONS.  
RX MEDLINE: 91282784.  
RA ISHIOKA C., SATO T., GAMOH M., SUZUKI T., SHIBATA H., KANAMARU R.,  
WAKI A., YAMAZAKI T.;  
BIOCHEM. BIOPHYS. RES. COMMUN. 177:901-906(1991).  
RN [23]  
RP VARIANTS OESOPHAGUS TUMORS L-152; A-155; H-175; F-176 AND H-273.  
RX MEDLINE: 91330175.  
RA CASSON A.G., MUKHOPADHYAY T., CLEARY K.R., RO J.Y., LEVIN B.,  
ROTH J.A.;  
CANCER RES. 51:4495-4499(1991).  
RN [24]  
RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN CHINA.  
RX MEDLINE: 91187113.  
RA HSU I.C., METCALF R.A., SUN T., WELSH J.A., WANG N.J., HARRIS C.C.;



RL NATURE 350:427-428(1991).  
 RN [25]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN SOUTH AFRICA.  
 RX MEDLINE: 91187114.  
 RA BRESSAC B., KEW M., WANDS J., OZTURK M.;  
 RL NATURE 350:429-431(1991).  
 RN [26]  
 RP VARIANTS IN ANOGENITAL CARCINOMAS.  
 RX MEDLINE: 93010989.  
 RA CROOK T., VOUSSDEN K.H.;  
 RL EMO J. 11:3935-3940(1992).  
 RN [27]  
 RP VARIANT WITH DUPLICATION OF PRO-177-HIS-178.  
 RX MEDLINE: 93265016.  
 RA BHATIA K., GUTIERREZ M.I., MAGRATH I.T.;  
 RL HUM. MOL. GENET. 1:207-208(1992).  
 RN [28]  
 RP VARIANTS IN BURKITT'S LYMPHOMAS.  
 RX MEDLINE: 93064692.  
 RA DUTH A., DEBOIRE B., ROMANO J.W., EHRHART J.C., FISCELLA M., MAY E.,  
 RA APPELLA E., MAY P.;  
 RL ONCOGENE 7:2161-2167(1992).  
 RN [29]  
 RP VARIANT NASOPHARYNGEAL CARCINOMA THR-280.  
 RX MEDLINE: 92335329.  
 RA SUN Y., HEGAMER G., HENG Y.-J., HILDSHEIM A., CHEN J.-Y., CAO Y.,  
 RA YAO K.-T., COLBURN N.H.;  
 RL PROC. NATL. ACADE. SCI. U.S.A. 89:6516-6520(1992).  
 RN [30]  
 RP VARIANTS IN COLON TUMORS.  
 RX MEDLINE: 93330562.  
 RA HAMELIN R., JEGO N., LAURENT-PUIG P., VIDAUD M., THOMAS G.;  
 RL ONCOGENE 8:2213-2220(1993).  
 RN [31]  
 RP CHARACTERIZATION OF VARIANT ALA-143.  
 RX MEDLINE: 94283378.  
 RA ZHANG W., GUO X.-Y., HU G.-Y., LIU W.-B., SHAY J.W., DEISSEROTH A.B.;  
 RL EMBO J. 13:2535-2544(1994).  
 RN [32]  
 RP VARIANTS LES HIS-175; ARG-193; GLN-248; CYS-273 AND TYR-275.  
 RX MEDLINE: 95195787.  
 RA FREIDBERG T., BARBIER N., YAN Y.-X., GARBER J.E., DREYFUS M.,  
 RA FREUMONT J.F., JR., LI F.P., FRIEND S.H.;  
 RL AM. J. HUM. GENET. 56:608-615(1995).  
 RN [33]  
 RP VARIANT LES HIS-175.  
 RX MEDLINE: 96423319.  
 RA VARLEY J.M., MCGROWN G., THORNCROFT M., TRICKER K.J., TEARE M.D.,  
 RA SANTIBANEZ-KOREF M.F., HOUUSTON R.S., MARTIN J., BIRCH J.M.,  
 RA EVANS D.G.R.;  
 RL J. MED. GENET. 32:942-945(1995).  
 RN [34]  
 RP VARIANTS BA PHE-176; SER-245; TRP-248; TRP-282 AND GLN-286.  
 RX MEDLINE: 96233927.  
 RA AUDEZEZ M.-P., ROBASKIEWICZ M., MERCIER B., NOUSBAUM J.-B.,  
 RA HARDY E., BAIL J.-P., VOLANT A., LOZACH P., GOEROU H., FEREC C.;  
 RL HUM. MUTAT. 7:109-113(1996).  
 Note: remainder of annotations omitted.

Query Match 100.0%; Score 86; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1,55e-08;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 127 SPALNMFOL 137  
 OY 1 SPALNMFOL 11

RESULT 3  
 ID P53-CERAE STANDARD; PRT; 393 AA.  
 AC P13481;  
 DT 01-JAN-1990 (REL. 13, CREATED)

DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS CERCOPIITHECUS AETHIOPS (GREEN MONKEY) (GRIVET).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 90045967.  
 RA RIGAUDY P., ECKHARDT W.;  
 RL NUCLEIC ACIDS RES. 17:8375-8375(1989).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND BCL-2 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC EMBL: X16384; G22796; -  
 CC PIR: S06594; S06594.  
 CC HSP: P04637; 10LG.  
 CC DR PROSITE: PS00348; P53: 1.  
 CC ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 CC KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 CC FT DOMAIN 1 68 ASP/GLU-RICH (ACIDIC).  
 CC FT DOMAIN 81 150 HYDROPHOBIC.  
 CC FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
 CC INTERACTION WITH DNA.  
 CC FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT MOD.RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
 CC SQ SEQUENCE 393 AA; 43696 MW; BBE7DC62 CRC32;

Query Match 95.3%; Score 82; DB 1; Length 393;  
 Best Local Similarity 90.9%; Pred. No. 1,84e-07;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 127 SPALNMFOL 137  
 OY 1 SPALNMFOL 11

RESULT 4  
 ID P53-HORSE STANDARD; PRT; 280 AA.  
 AC P79892; Q29481;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 DE TP53 OR P53.  
 OS EQUUS CABALLUS (HORSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PERISSODACTYLA.  
 RN [1]  
 RP SEQUENCE OF 1-263 FROM N.A.  
 RC TISSUE-SPLEEN;  
 RX MEDLINE: 97070350.  
 RA PAZZI K.A., KRAEGEL S.A., GRIFFEY S.M., THEON A.P., MADEWELL B.R.;  
 RL CANCER LETT. 107:125-130(1996).  
 RN [2]  
 RP SEQUENCE OF 76-280 FROM N.A.  
 RX MEDLINE: 96293865.  
 RA NASIR L., REID S.W.;  
 RL DNA SEQ. 6:185-187(1996).

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC EMBL: S83123; G1836145; -  
CC PROSITE: U37120; G1389675; -  
CC DR EMBL: P500348; P53; 1.  
CC ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
CC NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
CC NON\_TER 1 1  
CC DOMAIN 262 274 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC CONFLICT 79 79 T -> A (IN REF. 2).  
CC CONFLICT 83 83 L -> M (IN REF. 2).  
CC CONFLICT 111 111 A -> V (IN REF. 2).  
CC CONFLICT 138 138 G -> A (IN REF. 2).  
CC NON\_TER 280 280  
CC SEQUENCE 280 AA; 30985 MW; BA94F872 CRC32;  
Query Match 89.5%; Score 77; DB 1; Length 280;  
Best Local Similarity 81.8%; Pred. No. 3.79e-06;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 77 SPLINKFCOL 87  
1111111111  
1 SPALNKFCOL 11  
Yy  
RESULT 5  
ID P53\_SPEBE STANDARD; PRT: 314 AA.  
AC 064662;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53.  
OS SPERMOPHYTES BEECHYEI (BEECHY GROUND SQUIRREL).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-THYMUS;  
RX MEDLINE: 95007566.  
RA RIVKINA M.B.; CULLEN J.M.; ROBINSON W.S.; MARION P.L.;  
RL CANCER RES. 54:5430-5437(1994).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC EMBL: U43902; G1165312; -  
CC PROSITE: P500348; P53; 1.  
CC DR EMBL: P500348; P53; 1.  
CC ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
CC EUTHERIA; ARTIODACTYLA.

KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
FT NON\_TER 1 1  
FT DOMAIN 289 301 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT NON\_TER 314 314  
SO SEQUENCE 314 AA; 34618 MW; D07F433B CRC32;  
Query Match 89.5%; Score 77; DB 1; Length 314;  
Best Local Similarity 81.8%; Pred. No. 3.79e-06;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 105 SPLINKFCOL 115  
1111111111  
1 SPALNKFCOL 11  
Yy  
RESULT 6  
ID P53\_SHEEP STANDARD; PRT: 382 AA.  
AC P51664;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS OVIS ARIES (SHEEP).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; ARTIODACTYLA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD;  
RX MEDLINE: 95352828.  
RA DEQUIEDT F.; KETTMANN R.; BURNY A.; WILLEMS L.;  
RL DNA SEQ. 5:255-259(1995).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
CC EMBL: X81705; G602357; -  
CC DR PROSITE: P500348; P53; 1.  
CC ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 300 312 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 381 381 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 381 381 PHOSPHORYLATION (BY SIMILARITY).  
SO SEQUENCE 382 AA; 42809 MW; 0CB99A00 CRC32;  
Query Match 89.5%; Score 77; DB 1; Length 382;  
Best Local Similarity 81.8%; Pred. No. 3.79e-06;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 115 SPLINKFCOL 125  
1111111111  
1 SPALNKFCOL 11  
Yy  
RESULT 7  
ID P53\_BOVIN STANDARD; PRT: 386 AA.  
AC Q29628;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS BOS TAURUS (BOVINE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; ARTIODACTYLA.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 95352829.  
 RA DEQUEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
 RL DNA SEQ. 5:261-264(1995).  
 RN [2]  
 RP SEQUENCE OF 13-386 FROM N.A.  
 RC STRAIN-HOLSTEIN; TISSUE-THYMUS;  
 RX MEDLINE: 96401400.  
 RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;  
 RL VET. IMMUNOL. IMMUNOPATHOL. 52:53-63(1996).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: X81704; G602333; -  
 DR EMBL: D49825; G1729419; -  
 DR PROSITE: P500348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
 FT MOD\_RES 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 380 380 R -> T (IN REF. 2).  
 SQ SEQUENCE 386 AA; 43255 MW; 0322BF3D CRC32;  
 Query Match 89.5%; Score 77; DB 1; Length 386;  
 Best Local Similarity 81.8%; Pred. No. 3.79e-06;  
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 DB 119 SPLNKLFCOL 129  
 QY 1 SPALNKMFCOL 11  
 RESULT 8  
 ID P53\_CRIGR STANDARD; PRT: 393 AA.  
 AC 009185; G64397; P97258; P97788;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR P53.  
 OS CRICETUS GRISEUS (CHINESE HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CHANG W., MI L.J.;  
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 97183659.  
 RA LEE H., LARNER J.M., HAMLIN J.L.;  
 RL GENE 184:177-183(1997).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION

CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: Y08900; E303876; -  
 DR EMBL: Y08901; E303863; -  
 DR EMBL: U50395; G1842230; -  
 DR PROSITE: P500348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 74 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 75 150 HYDROPHOBIC.  
 FT DOMAIN 316 390 HIGHLY BASIC AND MAY BE INVOLVED IN  
 FT INTERACTION WITH DNA.  
 FT MOD\_RES 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
 FT VARIANT 133 133 L -> Q (IN CELL LINE V79-4).  
 FT VARIANT 135 135 C -> W (IN CELL LINE V79-4).  
 FT CONFLICT 103 103 Y -> F (IN REF. 2).  
 SQ SEQUENCE 393 AA; 43378 MW; 402EB149 CRC32;  
 Query Match 89.5%; Score 77; DB 1; Length 393;  
 Best Local Similarity 81.8%; Pred. No. 3.79e-06;  
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 DB 127 SPLNKLFCOL 137  
 QY 1 SPALNKMFCOL 11  
 RESULT 9  
 ID P53\_MESAU STANDARD; PRT: 396 AA.  
 AC 000366; P97276;  
 DT 01-DEC-1992 (REL. 24, CREATED)  
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS MESOCRICETUS AURATUS (GOLDEN HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-STRIAN; TISSUE-KIDNEY;  
 RX MEDLINE: 92210007;  
 RA LEGROS Y., MCINTYRE P., SOUSSI T.;  
 RL GENE 112:247-250(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA HOT E.W., WISEMAN R.;  
 RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: M75144; G191415; -

DR EMBL: U07182; G473579; -  
 DR PIR: JH0633;  
 DR HSSP: P04637; 1PES.  
 DR PROSITE: P500348; P53: 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 77 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 78 153 HYDROPHOBIC.  
 FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
 FT DOMAIN INTERACTION WITH DNA.  
 FT DOMAIN 314 326 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 395 395 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 188 188 G -> S (IN REF. 2).  
 SQ SEQUENCE 396 AA; 43631 MW; C2668ADE CRC32;  
 Query Match 89.5%; Score 77; DB 1; Length 396;  
 Best Local Similarity 81.8%; Fred. No. 3.79e-06;  
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Db 130 SPSINKLECOL 140  
 Qy 1 SPALNKFCOL 11  
 RESULT 10 STANDARD; PRT; 386 AA.  
 ID P53\_FELCA  
 AC P41685;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS FELIS SILVESTRIUS CATUS (CAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; CARNIVORA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LYMPH NODE;  
 RX MEDLINE: 94333960.  
 RA OKUDA M., UMEIDA A., SAKAI T., OHASHI T., MOMOI Y., YOUN H.Y.,  
 RA WATARI T., GOITSUKA R., TSUJIMOTO H., HASEGAWA A.;  
 RL INT. J. CANCER 58:602-607(1994).  
 RN [2]  
 RP SEQUENCE OF 34-354 FROM N.A.  
 RX MEDLINE: 94114699.  
 RA OKUDA M., UMEIDA A., MATSUMOTO Y., MOMOI Y., WATARI T., GOITSUKA R.,  
 RA O'BRIEN S.J., TSUJIMOTO H., HASEGAWA A.;  
 RL J. VET. MED. SCI. 55:801-805(1993).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: D26608; G538225; -  
 DR EMBL: D16460; G575528; -  
 DR PROSITE: P500348; P53: 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 285 285 K -> R (IN REF. 2).  
 SQ SEQUENCE 386 AA; 42692 MW; D6C7132A CRC32;

Query Match 88.4%; Score 76; DB 1; Length 386;  
 Best Local Similarity 81.8%; Fred. No. 6.89e-06;  
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Db 119 SPSINKLECOL 129  
 Qy 1 SPALNKFCOL 11  
 RESULT 11 STANDARD; PRT; 390 AA.  
 ID P53\_MOUSE  
 AC P02340;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR TRP53 OR P53.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 85027173.  
 RA BIENZ B., ZAKUT-HOURI R., GIVOL D., OREN M.;  
 RL EMBL J. 3:2179-2183(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84068204.  
 RA ZAKUT-HOURI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIVOL D.;  
 RL NATURE 306:594-597(1983).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84272240.  
 RA JENKINS J.R., RUDGE K., REDMOND S., WADE-EVANS A.;  
 RL NUCLEIC ACIDS RES. 12:5609-5626(1984).  
 RN [4]  
 RP SEQUENCE FROM N.A. (CLONES PCD53; P53-M11 AND P53-M8).  
 RX MEDLINE: 87064640.  
 RA ARAI N., NOMURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.,  
 RA ROTTER V.;  
 RL MOL. CELL. BIOL. 6:3232-3239(1986).  
 RN [5]  
 RP SEQUENCE OF 222-258 FROM N.A.  
 RX MEDLINE: 92115342.  
 RA BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BRENNER R.,  
 RA BALMAIN A.;  
 RL ONCOGENE 6:2363-2369(1991).  
 RN [6]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE: 86149247.  
 RA SAMAD A., ANDERSON C.W., CARROLL R.B.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:897-901(1986).  
 RN [7]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE: 91006019.  
 RA MEER D.W., SIMON S., KIRKMAN U., ECKHART W.;  
 RL EMBL J. 9:3253-3260(1990).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

EMBL: X00876; G871421; JOINED.  
 DR EMBL: X00877; G871421; JOINED.  
 DR EMBL: X00878; G871421; JOINED.  
 DR EMBL: X00879; G871421; JOINED.  
 DR EMBL: X00880; G871421; JOINED.  
 DR EMBL: X00881; G871421; JOINED.  
 DR EMBL: X00882; G871421; JOINED.  
 DR EMBL: X00883; G871421; JOINED.  
 DR EMBL: X00884; G871421; JOINED.  
 DR EMBL: X00885; G871421; JOINED.  
 DR EMBL: X01700; G200205; JOINED.  
 DR EMBL: X01237; G53576; -.  
 DR EMBL: X00741; G53571; -.  
 DR EMBL: M13872; G200199; -.  
 DR EMBL: M13873; G200201; -.  
 DR EMBL: M13874; G200203; ALT\_SEQ.  
 DR EMBL: S77930; G243255; -.  
 DR PIR: A02684; DNMS53.  
 DR PIR: A22739; A22739.  
 DR PIR: S38822; S38822.  
 DR HSSP: P04637; 1PES.  
 DR TRANSFAC: T01806; -.  
 DR MGD: MGI:98834; TRP53.  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS; DISEASE MUTATION.  
 FT DOMAIN 1 75 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 76 150 HYDROPHOBIC.  
 FT DOMAIN 276 390 HIGHLY BASIC AND MAY BE INVOLVED IN  
 INTERACTION WITH DNA.  
 FT DOMAIN 308 320 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 312 312 PHOSPHORYLATION.  
 FT MOD\_RES 389 389 PHOSPHORYLATION (BY CK2).  
 FT VARIANT 135 135 A -> V (CAN COOPERATE WITH AN ACTIVATED  
 RAS -> G (IN CLONE P53-M11).  
 FT VARIANT 168 168 E -> G (IN CLONE P53-M11).  
 FT CONFLICT 48 48 Q -> R (IN REF. 3).  
 FT CONFLICT 79 81 PVA -> QW (IN REF. 3).  
 SQ SEQUENCE 390 AA; 43458 MW; 8943DD93 CRC32;

Query Match 88.4%; Score 76; DB 1; Length 390;  
 Best Local Similarity 81.8%; Pred. No. 6.89e-06;  
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 124 SPUNKMFCOL 134  
 Oy 1 SPALKMFCOL 11  
 RESULT 12  
 ID P53\_XENLA STANDARD: PRT: 363 AA.  
 AC P07193;  
 DT 01-APR-1988 (REL. 07, CREATED)  
 DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE: 88143684.  
 RA SOUSST T., DE FROMENTEL C.C., MECHALI M., MAY P., KRESS M.;  
 RL ONCOGENE 1:71-78(1987).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE: 94134403.  
 RA HOEVER M., CLEMENT J.H., MEDLICH D., MONTENARH M., KNOCHEL W.;  
 RL ONCOGENE 9:109-120(1994).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A

TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 EXPRESSION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: UNBOUTHOUS.  
 CC -1- SIMILARITY: STRONG; TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: M36962; G214640; -.  
 DR EMBL: X05191; G64962; -.  
 DR EMBL: X77546; G468514; -.  
 DR EMBL: S68353; G545102; -.  
 DR PIR: A29376; A29376.  
 DR HSSP: P04637; 1PES.  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 281 293 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 362 362 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 52 52 T -> S (IN REF. 2).  
 FT CONFLICT 71 71 MISSING (IN REF. 2).  
 FT CONFLICT 296 296 MISSING (IN REF. 2).  
 SQ SEQUENCE 363 AA; 40692 MW; 75D7D796 CRC32;

Query Match 87.2%; Score 75; DB 1; Length 363;  
 Best Local Similarity 81.8%; Pred. No. 1.25e-05;  
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 101 SPUNKMFCOL 111  
 Oy 1 SPALKMFCOL 11  
 RESULT 13  
 ID P53\_BRARE STANDARD: PRT: 373 AA.  
 AC P79734;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS BRACHYDANTIO RERIO (ZEBRAFISH).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;  
 CC OSTEIFISHES; ACTINOPTERYGII; CYPRINIFORMES.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA BAILEY G.S., CHENG R., FORD B.L., O'NEAL P.E., MATHEWS C.Z.,  
 RA BRADFORD C.S., THONGTAN T., BARNES D.W., HENDRICKS J.D.;  
 RL MOL. MAR. BIOL. BIOTECHNOL. 6:88-97(1997).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 EXPRESSION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: STRONG; TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: U60804; G1778019; -.  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 280 296 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 372 372 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 373 AA; 41899 MW; 706A4B9C CRC32;

DB 95 SPDLNKLFCOL 105  
 QY 1 SPALNKMFCOL 11

RESULT 14  
 ID P53.SALIR STANDARD: PRT: 396 AA.  
 AC P5035:  
 DT 01-MAY-1992 (REL. 22, CREATED)  
 DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (TUMOR SUPPRESSING PROTEIN).  
 GN TP53.  
 OS SALMO IRIDEUS (RAINBOW TROUT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNAHOSTOMATA;  
 CC OSTEOICHTHYES; ACTINOPTERYGII; SALMONIFORMES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92210006.  
 RA DE FROMENTEL C.C., PADKEL F., CHAPUS A., BANAY C., MAY P., SOUSSI T.;  
 RL GENE 112:241-245(1992).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: M75145; G213829; -  
 DR PIR: JH0631; JH0631.  
 DR HSSP: P04637; 1PES.  
 DR PROSITE: P500348; P53. 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 302 318 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 395 395 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 396 AA; 43966 MW; BED1FC37 CRC32;

Query Match 87.28; Score 75; DB 1; Length 396;  
 Best Local Similarity 81.88; Pred. No. 1.25e-05;  
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 115 SPDLNKLFCOL 125  
 QY 1 SPALNKMFCOL 11

RESULT 15  
 ID P53.CANFA STANDARD: PRT: 276 AA.  
 AC 029537:  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN TP53.  
 OS CANIS FAMILIARIS (DOG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; CARNIVORA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BEAGLE:  
 RX MEDLINE: 95323915.  
 RA KRAESEL S.A., PAZZI K.A., MADEWELL B.R.;  
 RL CANCER LETT. 92:181-186(1995).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN

CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: S77819; G1000577; -  
 DR PROSITE: P500348; P53. 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT NON\_TER 1 1  
 FT DOMAIN 1 1  
 FT NON\_TER 1 1  
 FT NON\_TER 276 276 ASP/GLU-RICH (ACIDIC).  
 SQ SEQUENCE 276 AA; 30466 MW; 8C97AE44 CRC32;

Query Match 83.78; Score 72; DB 1; Length 276;  
 Best Local Similarity 81.88; Pred. No. 7.21e-05;  
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 90 SPDLNKLFCOL 100  
 QY 1 SPALNKMFCOL 11

Search completed: Fri Sep 11 13:33:26 1998  
 Job time : 8 secs.

\*\*\*\*\*

Release 3.1A John F. Collins, Bioinformatics Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Bioinformatics Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*

MPsrch\_p protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:33:44 1998; MasPar time 4.14 Seconds

Tabular output not generated. 111.756 Million cell updates/sec

Title: >US-08-452-843-15

Description: (1-11) from US08452843.pep

Perfect Score: 86

Sequence: 1 SPALNMFCOL 11

Scoring table: PAM 150

Gap 15

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

1:sp.fungi 2:sp.human 3:sp.invertebrate 4:sp.mammal  
5:sp.mhc 6:sp.organelle 7:sp.phage 8:sp.plant  
9:sp.bacteria 10:sp.proteob 11:sp.virus 12:sp.invertebrate  
13:sp.unclassified

Statistics: Mean 24.087; Variance 30.448; scale 0.791

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description            | Pred. No. |
|------------|-------|-------------|--------|----|--------|------------------------|-----------|
| 1          | 86    | 100.0       | 245    | 2  | 015085 | P53 TRANSFORMATION SUP | 8,83e-08  |
| 2          | 86    | 100.0       | 393    | 2  | 016809 | CELLULAR TUMOR ANTIGEN | 8,83e-08  |
| 3          | 86    | 100.0       | 393    | 2  | 016811 | CELLULAR TUMOR ANTIGEN | 8,83e-08  |
| 4          | 86    | 100.0       | 393    | 2  | 016807 | CELLULAR TUMOR ANTIGEN | 8,83e-08  |
| 5          | 86    | 100.0       | 393    | 2  | 016808 | CELLULAR TUMOR ANTIGEN | 8,83e-08  |
| 6          | 86    | 100.0       | 393    | 2  | 015087 | P53 TRANSFORMATION SUP | 8,83e-08  |
| 7          | 86    | 100.0       | 393    | 2  | 015088 | P53 TRANSFORMATION SUP | 8,83e-08  |
| 8          | 86    | 100.0       | 393    | 2  | 016810 | CELLULAR TUMOR ANTIGEN | 8,83e-08  |
| 9          | 86    | 100.0       | 393    | 2  | 016848 | CELLULAR TUMOR ANTIGEN | 8,83e-08  |
| 10         | 86    | 100.0       | 393    | 2  | 016535 | P53 TRANSFORMATION SUP | 8,83e-08  |
| 11         | 86    | 100.0       | 393    | 2  | 015086 | P53 TRANSFORMATION SUP | 8,83e-08  |
| 12         | 77    | 89.5        | 205    | 10 | 035873 | CELLULAR TUMOR ANTIGEN | 1.55e-05  |
| 13         | 77    | 89.5        | 238    | 11 | P89004 | P53 (FRAGMENT)         | 1.55e-05  |
| 14         | 77    | 89.5        | 238    | 11 | P89003 | P53 (FRAGMENT)         | 1.55e-05  |
| 15         | 77    | 89.5        | 238    | 11 | P90032 | P53 (FRAGMENT)         | 1.55e-05  |
| 16         | 77    | 89.5        | 378    | 11 | P89002 | P53 (FRAGMENT)         | 1.55e-05  |
| 17         | 77    | 89.5        | 391    | 13 | 036006 | CELLULAR TUMOR ANTIGEN | 1.55e-05  |
| 18         | 72    | 83.7        | 281    | 4  | 029475 | CELLULAR TUMOR ANTIGEN | 2.47e-04  |
| 19         | 72    | 83.7        | 285    | 4  | 095326 | CELLULAR TUMOR ANTIGEN | 2.47e-04  |
| 20         | 66    | 76.7        | 135    | 10 | 064451 | CELLULAR TUMOR ANTIGEN | 6.15e-03  |

|    |    |      |      |    |        |                        |          |
|----|----|------|------|----|--------|------------------------|----------|
| 21 | 61 | 70.9 | 1520 | 3  | 015829 | CARBANYL PHOSPHATE SYN | 8.07e-02 |
| 22 | 60 | 69.8 | 1466 | 3  | 023602 | ZR809.1.               | 1.33e-01 |
| 23 | 60 | 69.8 | 502  | 3  | 018301 | C29E6.3.               | 1.33e-01 |
| 24 | 58 | 67.4 | 499  | 2  | 015351 | P73 PROTEIN.           | 3.58e-01 |
| 25 | 58 | 67.4 | 636  | 2  | 015350 | P53-LIKE TRANSCRIPTION | 3.58e-01 |
| 26 | 55 | 64.0 | 360  | 3  | 020723 | FS3F4.7.               | 1.52e+00 |
| 27 | 54 | 62.8 | 344  | 3  | 025484 | GMP CYCLOHYDROLASE II/ | 2.44e+00 |
| 28 | 54 | 62.8 | 530  | 10 | 062789 | UDP-GLUCURONOSYLTRANSF | 2.44e+00 |
| 29 | 54 | 62.8 | 587  | 3  | 021432 | COSMD K11G12.          | 2.44e+00 |
| 30 | 53 | 61.6 | 167  | 3  | 094255 | COSMD K04A8.           | 3.89e+00 |
| 31 | 53 | 61.6 | 601  | 11 | 069309 | 67 KDA PROTEIN.        | 3.89e+00 |
| 32 | 53 | 61.6 | 623  | 11 | 067641 | U47.                   | 3.89e+00 |
| 33 | 53 | 61.6 | 634  | 10 | 035834 | KET PROTEIN (FRAGMENT) | 3.89e+00 |
| 34 | 53 | 61.6 | 4717 | 1  | 013676 | HYPOTHETICAL 337.8 KD  | 3.89e+00 |
| 35 | 52 | 60.5 | 435  | 9  | P77328 | SIMILAR TO B. SUBTILIS | 6.16e+00 |
| 36 | 51 | 59.3 | 56   | 9  | 049214 | HYPOTHETICAL PROTEIN ( | 9.70e+00 |
| 37 | 51 | 59.3 | 103  | 9  | 033157 | HYPOTHETICAL 11.2 KD P | 9.70e+00 |
| 38 | 51 | 59.3 | 390  | 9  | P77690 | FROM BASES 2360084 TO  | 9.70e+00 |
| 39 | 51 | 59.3 | 470  | 1  | 012049 | HYPOTHETICAL 53.7 KD P | 9.70e+00 |
| 40 | 51 | 59.3 | 474  | 11 | 040637 | PROTEASE.              | 9.70e+00 |
| 41 | 51 | 59.3 | 507  | 3  | 023038 | SIMILARITY TO C4-TYPE  | 9.70e+00 |
| 42 | 51 | 59.3 | 560  | 3  | 021388 | K09C8.3.               | 9.70e+00 |
| 43 | 51 | 59.3 | 2756 | 9  | 033904 | HYPOTHETICAL 292.7 KD  | 9.70e+00 |
| 44 | 50 | 58.1 | 340  | 11 | 069035 | G PROTEIN-COUPLED RECE | 1.52e+01 |
| 45 | 50 | 58.1 | 437  | 8  | 023024 | TIG11.14 PROTEIN.      | 1.52e+01 |

## ALIGNMENTS

|   |                                     |
|---|-------------------------------------|
| RESULT 1  | PRELIMINARY; PRT; 245 AA.           |
| ID 015085;  |                                     |
| AC 015085;  |                                     |
| DT 01-NOV-1996 (TREMBLREL. 01, CREATED)                           |                                     |
| DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)              |                                     |
| DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)            |                                     |
| DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).                      |                                     |
| GN P53.   |                                     |
| OS HOMO SAPIENS (HUMAN).  |                                     |
| OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; |                                     |
| OC EUTHERIA; PRIMATES.  |                                     |
| NC [1]  |                                     |
| RP SEQUENCE FROM N.A.   |                                     |
| RX MEDLINE: 92007731.   |                                     |
| RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;   |                                     |
| RL EMBL: X60010; G506433; -.                                      |                                     |
| DR EMBL: X60010; G506433; -.                                      |                                     |
| FT NON_TER 245  |                                     |
| FT SEQUENCE 245 AA; 27066 MW; 55B80C07 CRC32;                     |                                     |
| SO SEQUENCE   |                                     |
| Query Match   | 100.0%; Score 86; DB 2; Length 245; |
| Best Local Similarity 100.0%; Pred. No. 8,83e-08;                 |                                     |
| Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;       |                                     |
| DB 127 SPALNMFCOL 137   |                                     |
| QY 1 SPALNMFCOL 11  |                                     |
| RESULT 2  | PRELIMINARY; PRT; 393 AA.           |
| ID 016809;  |                                     |
| AC 016809;  |                                     |
| DT 01-NOV-1996 (TREMBLREL. 01, CREATED)                           |                                     |
| DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)              |                                     |
| DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)            |                                     |
| DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).                         |                                     |
| GN P53.   |                                     |
| OS HOMO SAPIENS (HUMAN).  |                                     |
| OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; |                                     |
| OC EUTHERIA; PRIMATES.  |                                     |
| NC [1]  |                                     |
| RP SEQUENCE FROM N.A.   |                                     |
| RX MEDLINE: 92007731.   |                                     |
| RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;   |                                     |

RL EMO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: X60019; G506451; .  
 DR PROSITE, PS00348; P53; 1.  
 DR ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 213 213 Q -> R.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43684 MW: CB70BD7F CRC32:  
 Query Match 100.0%; Score 86; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8.83e-08;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNMF001 137  
 Oy 1 SPALNMF001 11

RESULT 3  
 ID 016811 PRELIMINARY; PRT: 393 AA.  
 AC 016811;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 OS HOMO SAPIENS (HUMAN).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE: 85126934.  
 RA MALASHENSKI G., LAMB P., PIM D., PEACOCK J., CRANFORD L.,  
 RA BENCHIMOL S.;  
 RL EMO J. 3:3257-3262(1984).  
 CC [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE: 87064416.  
 RA LAMB P., CRANFORD L.;  
 RL MOL. CELL. BIOL. 6:1379-1385(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: M13121; G386994; .  
 DR EMBL: M13112; G386994; JOINED.  
 DR EMBL: M13113; G386994; JOINED.  
 DR EMBL: M13114; G386994; JOINED.  
 DR EMBL: M13115; G386994; JOINED.  
 DR EMBL: M13116; G386994; JOINED.  
 DR EMBL: M13117; G386994; JOINED.  
 DR EMBL: M13118; G386994; JOINED.  
 DR EMBL: M13119; G386994; JOINED.  
 DR EMBL: M13120; G386994; JOINED.  
 DR PROSITE, PS00348; P53; 1.  
 KW REPEAT: TUMOR ANTIGEN; ANTI-ONCOGENE: DNA-BINDING;  
 KW TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN;  
 KW PHOSPHORYLATION.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 3EA71431 CRC32:

Query Match 100.0%; Score 86; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8.83e-08;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNMF001 137  
 Oy 1 SPALNMF001 11

RESULT 4  
 ID 016807 PRELIMINARY; PRT: 393 AA.  
 AC 016807;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 OS HOMO SAPIENS (HUMAN).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: X60018; G506449; .  
 DR PROSITE, PS00348; P53; 1.  
 DR ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 SQ SEQUENCE 393 AA: 279BC9CB CRC32:

Query Match 100.0%; Score 86; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8.83e-08;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNMF001 137  
 Oy 1 SPALNMF001 11

RESULT 5  
 ID 016808 PRELIMINARY; PRT: 393 AA.  
 AC 016808;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 OS HOMO SAPIENS (HUMAN).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: X60018; G506449; .  
 DR PROSITE, PS00348; P53; 1.  
 DR ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.



FT VARIANT 163 163 H -> Y.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43627 MW: AFD8A9E3 CRC32:

Query Match  
 Best Local Similarity 100.0%; Score 86; DB 2; Length 393;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 127 SPALNKFQOL 137  
 |||||  
 QY 1 SPALNKFQOL 11

RESULT 6

ID Q15087 PRELIMINARY: PRT: 393 AA.

AC Q15087  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

OS P53. SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.

RN [1]  
 RP SEQUENCE FROM N.A.

RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60014; G506441; -

FT VARIANT 237 237 I -> M.  
 FT NON\_TER 393 393

SQ SEQUENCE 393 AA: 43694 MW: 9BB81992 CRC32:

Query Match  
 Best Local Similarity 100.0%; Score 86; DB 2; Length 393;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 127 SPALNKFQOL 137  
 |||||  
 QY 1 SPALNKFQOL 11

RESULT 7

ID Q15088 PRELIMINARY: PRT: 393 AA.

AC Q15088  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

OS P53. SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60016; G506445; -

FT VARIANT 238 238 Y -> C.  
 FT NON\_TER 393 393

SQ SEQUENCE 393 AA: 43713 MW: A01E1523 CRC32:

Query Match  
 Best Local Similarity 100.0%; Score 86; DB 2; Length 393;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 127 SPALNKFQOL 137  
 |||||  
 QY 1 SPALNKFQOL 11

RESULT 8  
 ID Q16810 PRELIMINARY: PRT: 393 AA.

AC Q16810  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).

OS P53. SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.

RN [1]  
 RP SEQUENCE FROM N.A.

RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBL: X60020; G506453; -

FT VARIANT 254 254 D -> V.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43714 MW: 5F914579 CRC32:

Query Match  
 Best Local Similarity 100.0%; Score 86; DB 2; Length 393;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 127 SPALNKFQOL 137  
 |||||  
 QY 1 SPALNKFQOL 11

RESULT 9

ID Q16848 PRELIMINARY: PRT: 393 AA.

AC Q16848  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.

OS P53. SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTER V.;  
 RL MOL. CELL. BIOL. 6:4650-4656(1986).

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: M14694; G339814; -  
 DR PROSITE: PS00348; P53; 1.  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ANTI-ONCOGENE; DNA-BINDING;  
 KW TRANSCRIPTION REGULATION; ACTIVATOR.

Query Match  
 Best Local Similarity 100.0%; Score 86; DB 2; Length 393;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNMFQOL 137

QY 1 SPALNMFQOL 11

RESULT 10 ID 016535 PRELIMINARY; PRT: 393 AA.

AC 016535;

DT 01-NOV-1996 (TREMBLREL, 01, CREATED)

DT 01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1996 (TREMBLREL, 01, LAST ANNOTATION UPDATE)

DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

GN P53

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RL MEDLINE: 92007731.

RA FARRRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;

RL EMBL: 10:2879-2887(1991).

DR EMBL: X60017; G506447; -

DR EMBL: X60015; G506443; -

FT VARIANT 248 248 Q -> R.

FT NON\_TER 393 393

SQ SEQUENCE 393 AA; 43684 MW; 239818A9 CRC32;

Query Match

Best Local Similarity 100.0%; Score 86; DB 2; Length 393;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNMFQOL 137

QY 1 SPALNMFQOL 11

RESULT 11 ID 015086 PRELIMINARY; PRT: 393 AA.

AC 015086;

DT 01-NOV-1996 (TREMBLREL, 01, CREATED)

DT 01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1996 (TREMBLREL, 01, LAST ANNOTATION UPDATE)

DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

GN P53

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RL MEDLINE: 92007731.

RA FARRRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;

RL EMBL: 10:2879-2887(1991).

DR EMBL: X60013; G506439; -

FT VARIANT 246 246 T -> M.

FT NON\_TER 393 393

SQ SEQUENCE 393 AA; 43682 MW; 943B62A3 CRC32;

Query Match

Best Local Similarity 100.0%; Score 86; DB 2; Length 393;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNMFQOL 137

QY 1 SPALNMFQOL 11

RESULT 12 ID 035873 PRELIMINARY; PRT: 205 AA.

AC 035873;

DT 01-JAN-1998 (TREMBLREL, 05, CREATED)

DT 01-JAN-1998 (TREMBLREL, 05, LAST SEQUENCE UPDATE)

DE 01-JAN-1998 (TREMBLREL, 05, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL, 05, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).

GN P53.

OS CRICETULUS GRISEUS (CHINESE HAMSTER).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; RODENTIA.

RN [1]

RP SEQUENCE FROM N.A.

RA RAINALDI G., MARCHETTI S., CAPECCHI B., MENEVERI R., PIRAS A.,

RA LEUZZI R.;

RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RA VATERONI L., MUSIO A., MENEVERI R., RAINALDI G.;

RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1 FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT

CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL

CC CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY

CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED

CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF

CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1 SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL: U74487; G2581764; -

DR PROSITE: P500348; P53: 1.

KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;

KW NUCLEAR PROTEIN; PHOSPHORYLATION.

FT NON\_TER 1 1

FT NON\_TER 205 205

SQ SEQUENCE 205 AA; 23122 MW; 680DDDDC CRC32;

Query Match

Best Local Similarity 81.8%; Score 77; DB 10; Length 205;

Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 2 SPALNMFQOL 12

QY 1 SPALNMFQOL 11

RESULT 13 ID P89004 PRELIMINARY; PRT: 238 AA.

AC P89004;

DT 01-MAY-1997 (TREMBLREL, 03, CREATED)

DT 01-MAY-1997 (TREMBLREL, 03, LAST SEQUENCE UPDATE)

DT 01-MAY-1997 (TREMBLREL, 03, LAST ANNOTATION UPDATE)

DE P53 (FRAGMENT).

OS MASTOMYS NATALENSIS PAPILLOMAVIRUS (MNPV).

OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-ECLOMA INDUCED BY LOXTIDINE;

RA LUQUE E.A., TANG L.H., MODLIN I.M.;

RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: U48618; G1813455; -

FT NON\_TER 1 1

SQ SEQUENCE 238 AA; 26704 MW; 097E01F9 CRC32;

Query Match

Best Local Similarity 81.8%; Score 77; DB 11; Length 238;

Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 20 SPALNMFQOL 30

QY 1 SPALNMFQOL 11

RESULT 14 ID P89003 PRELIMINARY; PRT: 286 AA.

AC P89003;

DT 01-MAY-1997 (TREMBLREL, 03, CREATED)

DT 01-MAY-1997 (TREMBLREL, 03, LAST SEQUENCE UPDATE)

DT 01-MAY-1997 (TREMBLREL, 03, LAST ANNOTATION UPDATE)

DE P53 (FRAGMENT).

OS. MASTOMYS NATALENSIS PAPILLOMAVIRUS (MNPV).  
 OC. VIRIDAE: DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.

RN [1]

RP. SEQUENCE FROM N.A.

RA. LUQUE E.A., TANG L.H., MODLIN I.M.;

RL. SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR. EMBL: U48617; G1813453; -

FT. NON\_TER 1

SO. SEQUENCE 286 AA; 32247 MW; 30F7C9FA CRC32;

Query Match

Best Local Similarity 89.5%; Score 77; DB 11; Length 286;

Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 20 SPSLNKLFCOL 30

QY 1 SPALNKMFCOL 11

RESULT 15

ID P90332 PRELIMINARY; PRT; 286 AA.

AC P90332;

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)

DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)

DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)

DE P53 (FRAGMENT)

OS. MASTOMYS NATALENSIS PAPILLOMAVIRUS (MNPV).

OC. VIRIDAE: DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.

RN [1]

RP. SEQUENCE FROM N.A.

RA. LUQUE E.A., TANG L.H., MODLIN I.M.;

RL. SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR. EMBL: U48619; G1813457; -

FT. NON\_TER 1

SO. SEQUENCE 286 AA; 32247 MW; 5B5D3CAD CRC32;

Query Match

Best Local Similarity 89.5%; Score 77; DB 11; Length 286;

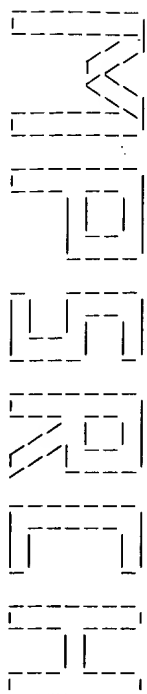
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 20 SPSLNKLFCOL 30

QY 1 SPALNKMFCOL 11

Search completed: Fri Sep 11 13:34:23 1998  
 Job time : 39 secs.

**THIS PAGE BLANK (USPTO)**



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mparch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:28:31 1998; Maspar time 2.63 Seconds  
61.575 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-14  
Description: (1-10) from US08452843.pep  
Perfect Score: 74  
Sequence: 1 APAPAPSMPL 10

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

a-geneseg32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 16.384; Variance 72.142; scale 0.227

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length DB | ID     | Description           | Pred. No. |
|------------|-------|-------------|-----------|--------|-----------------------|-----------|
| 1          | 74    | 100.0       | 253 24    | W28484 | Human p53 protein var | 3.94e+00  |
| 2          | 74    | 100.0       | 253 24    | W28483 | Human p53 protein var | 3.94e+00  |
| 3          | 74    | 100.0       | 270 24    | W28486 | Human p53 protein var | 3.94e+00  |
| 4          | 74    | 100.0       | 270 24    | W28485 | Human p53 protein var | 3.94e+00  |
| 5          | 74    | 100.0       | 319 24    | W28495 | Human p53 protein var | 3.94e+00  |
| 6          | 74    | 100.0       | 319 24    | W28496 | Human p53 protein var | 3.94e+00  |
| 7          | 74    | 100.0       | 335 24    | W28497 | Human p53 protein var | 3.94e+00  |
| 8          | 74    | 100.0       | 335 24    | W28498 | Human p53 protein var | 3.94e+00  |
| 9          | 74    | 100.0       | 353 24    | W28494 | Human p53 protein var | 3.94e+00  |
| 10         | 74    | 100.0       | 353 24    | W28493 | Human p53 protein var | 3.94e+00  |
| 11         | 74    | 100.0       | 363 21    | W13971 | Modified p53 variant  | 3.94e+00  |
| 12         | 74    | 100.0       | 363 21    | W13974 | Modified p53 variant  | 3.94e+00  |
| 13         | 74    | 100.0       | 363 21    | W13973 | Modified p53 variant  | 3.94e+00  |
| 14         | 74    | 100.0       | 363 21    | W13972 | Modified p53 variant  | 3.94e+00  |
| 15         | 74    | 100.0       | 363 24    | W28479 | Human p53 protein var | 3.94e+00  |
| 16         | 74    | 100.0       | 363 24    | W28480 | Human p53 protein var | 3.94e+00  |
| 17         | 74    | 100.0       | 363 21    | W13975 | Modified p53 variant  | 3.94e+00  |
| 18         | 74    | 100.0       | 363 21    | W13977 | Modified p53 variant  | 3.94e+00  |

|    |    |       |        |        |                       |          |
|----|----|-------|--------|--------|-----------------------|----------|
| 19 | 74 | 100.0 | 370 21 | W13957 | Chimeric p53 protein. | 3.94e+00 |
| 20 | 74 | 100.0 | 374 24 | W28481 | Human p53 protein var | 3.94e+00 |
| 21 | 74 | 100.0 | 374 24 | W28482 | Human p53 protein var | 3.94e+00 |
| 22 | 74 | 100.0 | 381 24 | W28490 | Human p53 protein var | 3.94e+00 |
| 23 | 74 | 100.0 | 381 24 | W28489 | Human p53 protein var | 3.94e+00 |
| 24 | 74 | 100.0 | 393 24 | W25155 | Human p53 variant fou | 3.94e+00 |
| 25 | 74 | 100.0 | 393 22 | W13953 | Human p53 variant     | 3.94e+00 |
| 26 | 74 | 100.0 | 393 22 | W13948 | Human p53 variant     | 3.94e+00 |
| 27 | 74 | 100.0 | 393 21 | W05345 | Human p53 mutant N239 | 3.94e+00 |
| 28 | 74 | 100.0 | 393 22 | W13951 | Human tumour-derived  | 3.94e+00 |
| 29 | 74 | 100.0 | 393 22 | W13952 | Human tumour-derived  | 3.94e+00 |
| 30 | 74 | 100.0 | 393 22 | W13952 | Human tumour-derived  | 3.94e+00 |
| 31 | 74 | 100.0 | 393 18 | R91918 | Wild type p53 protein | 3.94e+00 |
| 32 | 74 | 100.0 | 393 18 | W02617 | Human p53 tumour supp | 3.94e+00 |
| 33 | 74 | 100.0 | 393 21 | W05348 | Human p53 mutant R282 | 3.94e+00 |
| 34 | 74 | 100.0 | 393 21 | W05344 | Human p53             | 3.94e+00 |
| 35 | 74 | 100.0 | 393 21 | W13968 | Modified p53 variant  | 3.94e+00 |
| 36 | 74 | 100.0 | 393 21 | W13970 | Modified p53 variant  | 3.94e+00 |
| 37 | 74 | 100.0 | 393 21 | W13969 | Modified p53 variant  | 3.94e+00 |
| 38 | 74 | 100.0 | 401 24 | W28487 | Human p53 protein var | 3.94e+00 |
| 39 | 74 | 100.0 | 401 24 | W28488 | Human p53 protein var | 3.94e+00 |
| 40 | 74 | 100.0 | 406 21 | W13964 | Chimeric p53 protein. | 3.94e+00 |
| 41 | 74 | 100.0 | 406 21 | W13966 | Chimeric p53 protein. | 3.94e+00 |
| 42 | 74 | 100.0 | 411 21 | W13967 | Chimeric p53 protein. | 3.94e+00 |
| 43 | 74 | 100.0 | 533 23 | W19763 | p53-GM-CSF immunostim | 3.94e+00 |
| 44 | 74 | 100.0 | 535 24 | W28492 | Human p53 protein var | 3.94e+00 |
| 45 | 74 | 100.0 | 535 24 | W28491 | Human p53 protein var | 3.94e+00 |

## ALIGNMENTS

RESULT 1  
AC W28484 standard; Protein: 253 AA.  
ID W28484;  
DE 25-NOV-1997 (first entry)  
DE Human p53 protein variant V-367H.  
KW Leucine zipper domain; LZD; Oligomerisation domain; mutant; mutin;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
OS tumour suppression; apoptosis.  
OS Chimeric - Homo sapiens.  
OS Chimeric - Herpes simplex virus.  
OS Synthetic.  
FH Key  
FT Location/Qualifiers  
FT misc\_difference 189  
FT /note= "Arg residue at position 182 of wild-type  
FT p53 has been mutated to His"  
PN  
PD W09704092-A1.  
PF 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E.  
DR MPI: 97-132633/12.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 32; Page -; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the p53  
CC transactivating domain (amino acids 1-74) deleted and replaced by  
CC the transactivating domain (TD) from herpes simplex virus viral  
CC protein VP16 (amino acids 411-490). The present sequence is that of  
CC a specifically claimed p53 variant designated V-367 and comprising  
CC the VP16 TD and amino acids 75-367 of human wild-type p53 (but with  
CC Arg182 replaced by His). The p53 variants are more active and more  
CC stable tumour suppressors and apoptosis-inducing agents than wild-type  
CC p53 and are active where the wild-type protein is not.  
CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant V-367).  
SO Sequence 253 AA;

Query Match 100.0%; Score 74; DB 24; Length 253;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 92 apapapswpl 101

1 APAPAPSWPL 10

RESULT 2

ID W28483 standard: Protein; 253 AA.

AC W28483; 1997 (first entry)

DE Human p53 protein variant V-367 encoded by PEC141.

KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;

KW substitution; replacement; transactivation; viral protein VP16; HSV;

KW anti-oncogene; hyperproliferation; cancer; restenosis;

KW tumour suppression; apoptosis.

OS Chimeric - Homo sapiens.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

Location/Qualifiers  
misc\_difference 189 /note="Arg residue at position 182 of wild-type p53 has been mutated to His"

MO9704092-A1.

06-FEB-1997.

17-JUL-1996; F01111.

19-JUL-1995; FR-008729.

(RHON ) RHONE POULENC RORER SA.

Bracco L, Conseiller E;

Query Match 100.0%; Score 74; DB 24; Length 253;

Best Local Similarity 100.0%; Pred. No. 3.94e+00;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 92 apapapswpl 101

1 APAPAPSWPL 10

RESULT 3

ID W28486 standard: Protein; 270 AA.

AC W28486; 1997 (first entry)

DE Human p53 protein variant V-ASH.

KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;

KW substitution; replacement; transactivation; viral protein VP16; HSV;

KW anti-oncogene; hyperproliferation; cancer; restenosis; murine;

KW tumour suppression; apoptosis; alternative splicing; AS form.

OS Chimeric - Homo sapiens.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

WPI: 97-132633/12.

PT New p53 variants e.g. with oligomerisation domain replaced by

PT leucine zipper useful for treating hyper-proliferative disorders,

PT esp. cancer and restenosis

PS Claim 33; Page -; 133pp; French.

CC Claimed variants of protein p53 have at least part of the p53

CC transactivation domain (amino acids 1-74) deleted and replaced by

CC the transactivating domain (TTD) from herpes simplex virus viral

CC protein VP16 (amino acids 411-490). The present sequence is that of

CC a specifically claimed p53 variant designated V-ASH and comprising

CC the VP16 TD with amino acids 75-366 of human wild-type p53 (but

CC with Arg182 replaced by His), followed by the last 19 C-terminal

CC amino acids of the alternatively spliced (AS) form of murine p53

CC (encoded by a synthetic linker). The variants are more active and

CC more stable tumour suppressors and apoptosis-inducing agents than

CC wild-type p53 and are active where the wild-type protein is not.

CC (Note: this sequence does not appear in the specification and has

CC been produced by modifying the given sequence of variant V-AS).

Sequence 270 AA;

Query Match 100.0%; Score 74; DB 24; Length 270;

Best Local Similarity 100.0%; Pred. No. 3.94e+00;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 92 apapapswpl 101

1 APAPAPSWPL 10

RESULT 4

ID W28485 standard: Protein; 270 AA.

AC W28485; 1997 (first entry)

DE Human p53 protein variant V-AS encoded by PEC143.

KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;

KW substitution; replacement; transactivation; viral protein VP16; HSV;

KW anti-oncogene; hyperproliferation; cancer; restenosis; murine;

KW tumour suppression; apoptosis; alternative splicing; AS form.

OS Chimeric - Homo sapiens.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

OS Chimeric - Herpes simplex virus.

OS Synthetic.

Location/Qualifiers  
misc\_difference 189 /note="Arg residue at position 182 of wild-type p53 has been mutated to His"

MO9704092-A1.

06-FEB-1997.

17-JUL-1996; F01111.

19-JUL-1995; FR-008729.

(RHON ) RHONE POULENC RORER SA.

Bracco L, Conseiller E;

1 APAPAPSWPL 10

RESULT 5  
ID W28495 standard; Protein: 319 AA.  
AC W28495.  
DE 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360-325 encoded by PEC178.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
PN WO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PR 19-JUL-1995; FR-008729.  
PI Bracco L, Conseiller E.  
PI Bracco L, Conseiller E.  
PI WPI: 97-132633/12.  
DR N-PSDB: T86223.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 38; Pages 92-94; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360-325 and comprising  
CC the 325-360 domain, amino acids 75-325 of human wild-type p53 and a  
CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).  
SQ Sequence 319 AA.

Query Match 100.0%; Score 74; DB 24; Length 319;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 48 apapapswpl 57  
OY 1 APAPAPSWPL 10  
IIIIIIIIII

RESULT 6  
ID W28496 standard; Protein: 319 AA.  
AC W28496.  
DE 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360-325H.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
PN WO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PR 19-JUL-1995; FR-008729.  
PI Bracco L, Conseiller E.  
PI Bracco L, Conseiller E.  
PI WPI: 97-132633/12.  
DR N-PSDB: T86223.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 39; Pages 94-95; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360h-325 and comprising  
CC the 325-360 domain, separated from amino acids 75-325 of human  
CC wild-type p53 by a synthetic hinge sequence (GlySer)<sub>3</sub>, and with a  
CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).  
SQ Sequence 335 AA.

Query Match 100.0%; Score 74; DB 24; Length 335;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PT esp. cancer and restenosis  
PS Claim 38; Page 7; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360-325H and comprising  
CC the 325-360 domain, amino acids 75-325 of human wild-type p53 (but with  
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
CC The p53 variants are more active and more stable tumour suppressors  
CC and apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not, i.e. they are not inactivated by dominant  
CC negative or oncogenic mutants, nor by other cellular proteins (because  
CC the leucine zipper domain prevents formation of inactive mixed  
CC oligomers).  
CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant 360-325).  
SQ Sequence 319 AA.

Query Match 100.0%; Score 74; DB 24; Length 319;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 48 apapapswpl 57  
OY 1 APAPAPSWPL 10  
IIIIIIIIII

RESULT 7  
ID W28497 standard; Protein: 335 AA.  
AC W28497.  
DE 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360h-325 encoded by PEC179.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
KW substitution; replacement; transactivation; hinge region;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
PN WO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PR 19-JUL-1995; FR-008729.  
PI Bracco L, Conseiller E.  
PI Bracco L, Conseiller E.  
PI WPI: 97-132633/12.  
DR N-PSDB: T86224.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 39; Pages 94-95; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360h-325 and comprising  
CC the 325-360 domain, separated from amino acids 75-325 of human  
CC wild-type p53 by a synthetic hinge sequence (GlySer)<sub>3</sub>, and with a  
CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).  
SQ Sequence 335 AA.

Query Match 100.0%; Score 74; DB 24; Length 335;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 48 apapapswpl 57  
OY 1 APAPAPSWPL 10  
IIIIIIIIII

RESULT 8  
ID W28498 standard; Protein: 335 AA.  
AC W28498.  
DE 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360h-325 encoded by PEC179.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
KW substitution; replacement; transactivation; hinge region;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
PN WO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PR 19-JUL-1995; FR-008729.  
PI Bracco L, Conseiller E.  
PI Bracco L, Conseiller E.  
PI WPI: 97-132633/12.  
DR N-PSDB: T86224.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 39; Pages 94-95; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360h-325 and comprising  
CC the 325-360 domain, separated from amino acids 75-325 of human  
CC wild-type p53 by a synthetic hinge sequence (GlySer)<sub>3</sub>, and with a  
CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).  
SQ Sequence 335 AA.

Query Match 100.0%; Score 74; DB 24; Length 335;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 64 apapapswpl 73  
| | | | |  
OY 1 APAPAPSWPL 10

## RESULT 8

ID W28498 standard; Protein: 335 AA.  
AC W28498;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360h-325H.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutetin;  
KW substitution; replacement; transactivation; hinge region;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
FH Key location/Qualifiers  
FT region 39..53  
FT /label= hinge  
FT misc-difference 161  
FT /note= "Arg residue at position 182 of wild-type  
p53 has been mutated to His"  
PN MO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR WPI: 97-132633/12.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
leucine zipper useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 39; Page -; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360h-325H and comprising  
CC the 325-360 domain, separated from amino acids 75-325 of human  
CC wild-type p53 (but with Arg182 replaced by His) by a synthetic hinge  
CC sequence (Gly4Ser)3, and with a leucine zipper domain at the C-terminal.  
CC The p53 variants are more active and more stable tumour suppressors  
CC and apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not, i.e. they are not inactivated by dominant  
CC negative or oncogenic mutants, nor by other cellular proteins (because  
CC the leucine zipper domain prevents formation of inactive mixed  
CC oligomers).  
CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant 360h-325).  
SQ Sequence 335 AA;

Query Match 100.0%; Score 74; DB 24; Length 335;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 64 apapapswpl 73  
| | | | |  
OY 1 APAPAPSWPL 10

## RESULT 9

ID W28494 standard; Protein: 353 AA.  
AC W28494;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 393-325H.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutetin;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.

OS Synthetic.

FH Key location/Qualifiers

FT misc-difference 179 /note= "Arg residue at position 182 of wild-type

FT p53 has been mutated to His"

PN MO9704092-A1.

PD 06-FEB-1997.

PF 17-JUL-1996; F01111.

PR 19-JUL-1995; FR-008729.

PA (RHON ) RHONE POULENC RORER SA.

PI Bracco L, Conseiller E;

DR WPI: 97-132633/12.

PT New p53 variants e.g. with oligomerisation domain replaced by

PT leucine zipper - useful for treating hyper-proliferative disorders,

PT esp. cancer and restenosis

PS Claim 39; Page -; 133pp; French.

CC Claimed variants of protein p53 have at least part of the

CC oligomerisation domain deleted and replaced by a leucine zipper

CC domain. The mutants preferably also have at least part of the p53

CC transactivation domain (amino acids 1-74) deleted and replaced by

CC the domain 325-393 of p53. The present sequence is that of a

CC specifically claimed p53 variant designated 393-325H and comprising

CC the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with

CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.

CC The p53 variants are more active and more stable tumour suppressors

CC and apoptosis-inducing agents than wild-type p53 and are active where

CC the wild-type protein is not, i.e. they are not inactivated by dominant

CC negative or oncogenic mutants, nor by other cellular proteins (because

CC the leucine zipper domain prevents formation of inactive mixed

CC oligomers).

CC (Note: this sequence does not appear in the specification and has

CC been produced by modifying the given sequence of variant 393-325).

SQ Sequence 353 AA;

Query Match 100.0%; Score 74; DB 24; Length 353;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 82 apapapswpl 91  
| | | | |  
OY 1 APAPAPSWPL 10

## RESULT 10

ID W28493 standard; Protein: 353 AA.  
AC W28493;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 393-325 encoded by PEC177.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutetin;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
PN MO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR WPI: 97-132633/12.  
DR N-PSDB; T86222.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 39; Pages 90-92; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-393 of p53. The present sequence is that of  
CC a specifically claimed p53 variant designated 393-325 and comprising  
CC the 325-393 domain, amino acids 75-325 of human wild-type p53 and a



CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).  
SQ Sequence 353 AA;

Db 82 apapapswpl 91  
| | | | | | | | | |  
QY 1 APAPAPSWPL 10

Query Match 100.0%; Score 74; DB 24; Length 353;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11  
ID W13971 standard; Protein: 363 AA.

AC W13971;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53R284del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KM apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazoneis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer.  
PS Example 1; 51-52; 82pp; English.  
CC Modified p53 variant p53R284del364-393 (W13971) has a Thr284 to Arg  
CC substitution (see also W13949) and a deletion of the C-terminal 30  
CC amino acids. The T284R substitution, introduced by site-directed  
CC mutagenesis of p53 DNA, provides a novel p53-DNA contact between a  
CC phosphate of the DNA backbone and p53. The C-terminal deletion  
CC permits in vitro DNA binding. The variant provides the means for  
CC pharmacological rescue of p53 function in cancer patients. Other  
CC modified p53 constructs (W13949-50, W13953-54, W13968-77) have also  
CC been produced. Nucleic acids coding for modified p53 can be used  
CC for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 74; DB 21; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 apapapswpl 93  
| | | | | | | | | |  
QY 1 APAPAPSWPL 10

RESULT 12  
ID W13974 standard; Protein: 363 AA.  
AC W13974;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53H273del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KM apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazoneis TD;  
DR WPI; 97-202618/18.

PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer.  
PS Example 1; 56-57; 82pp; English.

CC Modified p53 variant p53H273del364-393 (W13974) has the tumour-  
CC derived histidine 273 mutation (see also W13952) and a deletion  
CC of the C-terminal 30 amino acids of wild-type p53 (see also  
CC W13948). His273 is a Class I p53 tumour mutation that affects DNA  
CC binding. The C-terminal deletion, introduced by site-directed  
CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
CC tumour mutant. This provides the means for pharmacological rescue  
CC of p53 function in cancer patients. Other modified p53 constructs  
CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
CC acids coding for modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 74; DB 21; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 apapapswpl 93  
| | | | | | | | | |  
QY 1 APAPAPSWPL 10

RESULT 13  
ID W13973 standard; Protein: 363 AA.

AC W13973;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53Q248R284del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KM apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazoneis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer.  
PS Example 1; 54-56; 82pp; English.  
CC Modified p53 variant p53Q248R284del364-393 (W13973) has the tumour-  
CC derived Gln248 mutation (see also W13951), a Thr284 to Arg substn.  
CC (see also W13949) and a deletion of the 30 C-terminal amino acids  
CC of wild-type p53 (W13948). Gln248 is a Class I p53 tumour mutation  
CC that affects DNA binding. The T284R substitution, introduced by  
CC site-directed mutagenesis of p53 DNA, provides a novel p53-DNA  
CC contact between a phosphate of the DNA backbone and p53, and  
CC restores DNA binding. The C-terminal deletion permits in vitro  
CC DNA binding. The construct provides the means for pharmacological  
CC rescue of p53 function in cancer patients. Other modified p53  
CC constructs (W13949-50, W13953-54, W13968-77) have also been  
CC produced. Nucleic acids coding for modified p53 can be used for  
CC cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 74; DB 21; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.94e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 apapapswpl 93  
| | | | | | | | | |  
QY 1 APAPAPSWPL 10

RESULT 14  
ID W13972 standard; Protein: 363 AA.  
AC W13972;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53Q248del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;

KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD:  
 DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 treatment of cancer.  
 PS Example 1; 53-54; 82pp; English.  
 CC Modified p53 variant p53Q248del1364-393 (W13972) has the tumour-  
 derived glutamine 248 mutation (see also W13951) and a deletion  
 of the C-terminal 30 amino acids of wild-type p53 (see also  
 W13948). Gln248 is a Class I p53 tumour mutation that affects DNA  
 binding. The C-terminal deletion, introduced by site-directed  
 mutagenesis of p53 DNA, activates the DNA binding of the p53  
 tumour mutant. This provides the means for pharmacological rescue  
 of p53 function in cancer patients. Other modified p53 constructs  
 (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
 acids coding for modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 74; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 3.94e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 apapapswp1 93  
 1 APAPAPSWPL 10

RESULT 15  
 ID W28479 standard; Protein; 363 AA.  
 AC W28479.  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-325 encoded by pEC114.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON) RHONE-POULENC RORER SA.  
 PI BRACCO L. Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB: T86215.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 leucine zipper - useful for treating hyper-proliferative disorders,  
 esp. cancer and restenosis  
 PT esp. cancer and restenosis  
 PS Claim 30. Pages 76-78; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 oligomerisation domain deleted and replaced by a leucine zipper  
 domain. The mutants, preferably also have at least part of the p53  
 transactivation domain (amino acids 1-74) deleted and replaced by  
 the transactivating domain (TD) from herpes simplex virus viral  
 protein VP16 (amino acids 411-490). The present sequence is that of  
 a specifically claimed p53 variant designated V-325 and comprising  
 the VP16 TD, amino acids 75-325 of human wild-type p53 and a  
 leucine zipper domain at the C-terminal. The p53 variants are  
 more active and more stable tumour suppressors and apoptosis-inducing  
 agents than wild-type p53 and are active where the wild-type protein  
 is not, i.e. they are not inactivated by dominant negative or oncogenic  
 mutants, nor by other cellular proteins (because the leucine zipper  
 domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 74; DB 24; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 3.94e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 92 apapapswp1 101  
 1 APAPAPSWPL 10

Search completed: Fri Sep 11 13:28:41 1998  
 Job time : 10 secs.

(TM)

Distribution rights by Oxford Molecular Ltd

Run on: Fri Sep 11 13:28:59 1998; MasPar time 3.04 Seconds

120.097 Million cell updates/sec

1 APAPAPSWPL 10

Gap 15

Post-processing: Minimum Match 0%  
Listing first 45 summaries

```
1:plr1 2:plr2 3:plr3 4:plr4 5:nr13d
```

Mean 22.554; Variance 45.898; scale 0.491

and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No | Score | Query Match | Length | DB | ID      | Description            | Pred. No. |
|-----------|-------|-------------|--------|----|---------|------------------------|-----------|
| 1         | 74    | 100.0       | 393    | 1  | DNRHS3  | cellular tumor antigen | 3.27e-02  |
| 2         | 74    | 100.0       | 393    | 2  | S06594  | cellular tumor antigen | 3.27e-02  |
| 3         | 66    | 89.2        | 391    | 2  | J51648  | tumor suppressor p53   | 5.50e-07  |
| 4         | 64    | 86.5        | 386    | 2  | SJ6193  | cellular tumor antigen | 1.09e+00  |
| 5         | 63    | 85.1        | 381    | 2  | S38824  | cellular tumor antigen | 1.53e+00  |
| 6         | 63    | 85.1        | 390    | 1  | DNRHS3  | cellular tumor antigen | 1.53e+00  |
| 7         | 56    | 75.7        | 393    | 2  | S02192  | cellular tumor antigen | 1.54e+01  |
| 8         | 56    | 75.7        | 393    | 2  | J06176  | tumor suppressor prot  | 1.54e+01  |
| 9         | 56    | 75.7        | 386    | 2  | JH0633  | cellular tumor antigen | 1.54e+01  |
| 10        | 56    | 75.7        | 564    | 2  | B43776  | drebrin E1 - chicken   | 1.54e+01  |
| 11        | 56    | 75.7        | 593    | 2  | I51213  | drebrin - chicken (fr  | 1.54e+01  |
| 12        | 56    | 75.7        | 607    | 2  | A43776  | drebrin E2 - chicken   | 1.54e+01  |
| 13        | 56    | 75.7        | 837    | 2  | C69187  | conserved hypothetical | 1.54e+01  |
| 14        | 56    | 75.7        | 2554   | 1  | TYEFL7L | kinase-related protei  | 1.54e+01  |
| 15        | 55    | 74.3        | 27     | 2  | B39690  | neurotoxin V-4 - bark  | 2.11e+01  |
| 16        | 55    | 74.3        | 66     | 2  | B23727  | disintegrin and metal  | 2.11e+01  |
| 17        | 55    | 74.3        | 814    | 2  | S03390  | endothelin 1 precursor | 2.11e+01  |
| 18        | 54    | 73.0        | 212    | 1  | ANHU1   | kappa-type opioid rec  | 2.90e+01  |
| 19        | 54    | 73.0        | 288    | 2  | A43083  | homeotic protein cdx-  | 2.90e+01  |
| 20        | 54    | 73.0        | 440    | 2  | A44081  | homeotic protein cdx-  | 3.96e+01  |
| 21        | 53    | 71.6        | 311    | 2  | A53808  | homeotic protein cdx-  | 3.96e+01  |
| 22        | 53    | 71.6        | 440    | 2  | S63358  | familial Alzheimer's   | 3.96e+01  |
| 23        | 53    | 71.6        | 495    | 2  | S53179  | hypothetical protein   | 3.96e+01  |

|    |    |      |      |   |         |                       |          |
|----|----|------|------|---|---------|-----------------------|----------|
| 45 | 49 | 66.2 | 1742 | 2 | S24600  | projectin - fruit fly | 3.96e+01 |
| 44 | 49 | 66.2 | 1247 | 2 | A33812  | inophthoroseceptor re | 1.34e+02 |
| 43 | 49 | 66.2 | 852  | 2 | S38415  | guanine nucleotide re | 1.34e+02 |
| 42 | 49 | 66.2 | 576  | 2 | S12792  | protein-tyrosine kina | 1.34e+02 |
| 41 | 49 | 66.2 | 503  | 2 | S55589  | D-nopalin dehydrogen  | 1.34e+02 |
| 40 | 49 | 66.2 | 245  | 2 | C64406  | N5-methyl-tetrahydro  | 1.34e+02 |
| 39 | 49 | 66.2 | 41   | 2 | A42064  | lactam utilization pr | 1.34e+02 |
| 38 | 50 | 67.6 | 1839 | 1 | RRWPEM  | RNA-directed RNA poly | 9.94e+01 |
| 37 | 50 | 67.6 | 1287 | 2 | S55954  | viral mRNA translatio | 9.94e+01 |
| 36 | 50 | 67.6 | 415  | 2 | A54550  | probable transposase  | 9.94e+01 |
| 35 | 50 | 67.6 | 333  | 2 | JC4695  | glycogenin glucosyltr | 9.94e+01 |
| 34 | 50 | 67.6 | 332  | 2 | A45094  | glycogenin glucosyltr | 9.94e+01 |
| 33 | 50 | 67.6 | 279  | 2 | S45141  | glycoethel protein    | 9.94e+01 |
| 32 | 50 | 67.6 | 204  | 2 | S18657  | hypothetical protein  | 9.94e+01 |
| 31 | 51 | 68.9 | 2411 | 2 | A48266  | protein-tyrosine kina | 7.34e+01 |
| 30 | 51 | 68.9 | 864  | 2 | A82866  | protein-tyrosine kina | 7.34e+01 |
| 29 | 51 | 68.9 | 550  | 1 | VGEBE18 | glycoprotein E - huma | 7.34e+01 |
| 28 | 51 | 68.9 | 472  | 2 | S36548  | I2 protein - human pa | 7.34e+01 |
| 27 | 51 | 68.9 | 464  | 2 | A33675  | protein-tyrosine kina | 7.34e+01 |
| 26 | 51 | 68.9 | 291  | 2 | E21844  | spdb protein - Strept | 7.34e+01 |
| 25 | 52 | 70.3 | 605  | 2 | I38755  | neuron-restrictive si | 5.40e+01 |
| 24 | 53 | 71.6 | 681  | 2 | A86655  | methylinomyI-CoA car  | 3.96e+01 |

## ALIGNMENTS

|                   |   |  |                |
|-------------------|---|--|----------------|
| RESULT            | 1 | DNHU53   | #type complete |
| ENTRY             |   | cellular tumor antigen p53 - human   |                |
| TITLE             |   | cellular phosphoprotein p53; oncoprotein p53; transformation suppressor p53; tumor suppressor p53  |                |
| ALTERNATE_NAMES   |   | #formal_name Homo sapiens #common_name man   |                |
| ORGANISM          |   | 05-oct-1988 #sequence_revision 18-Nov-1994 #text_change 18-SEP-1997  |                |
| DATE              |   |  |                |
| ACCESSIONS        |   | A25224; A43073; JT0436; S40773; S42669; A22837; A55060; A25397; B25397; S42452; S42453; I38082; I38083; I38084; I38085; I38086; I38087; I38088; I38089; I38090; I38091; I38092; I38093; A44905; I58354; I78850; S60153 |                |
| REFERENCE         |   |  |                |
| #authors          |   | Lamb, P.; Crawford, L.   |                |
| #journal          |   | Mol. Cell. Biol. (1986) 6:1379-1385  |                |
| #title            |   | Characterization of the human p53 gene.  |                |
| #cross-references |   | Characterization of the human p53 gene.  |                |
| #accession        |   | Accession M01D:87064416  |                |
|                   |   | A25224   |                |
|                   |   | ##molecule_type DNA  |                |
|                   |   | ##residues 1-393 ##label LAM   |                |
|                   |   | #cross-references EMBL:X01405; GB:MI3121; GB:N00032; NID:q189460; PID:g386694  |                |
| REFERENCE         |   |  |                |
| #authors          |   | JT0436   |                |
|                   |   | Buchman, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P. Georgiev, G.P.  |                |
| #journal          |   | Gene (1988) 70:245-252   |                |
| #title            |   | A variation in the structure of the protein-coding region of the human p53 gene.   |                |
| #cross-references |   | the human p53 gene.  |                |
| #accession        |   | Accession M01D:89108008  |                |
|                   |   | A43073   |                |
|                   |   | ##molecule_type DNA  |                |
|                   |   | ##residues 1-393 ##label BUC   |                |
|                   |   | ##note this 72-Arg allele appears to be about 5 times more frequent than the 72-Pro allele   |                |
| #accession        |   | JT0436   |                |
|                   |   | ##molecule_type DNA  |                |
|                   |   | ##residues 1-771, 'P', 73-393 ##label BU2  |                |
|                   |   | #cross-references EMBL:M22898; NID:9189474; PID:g189476  |                |
|                   |   | ##note this 72-Pro allele was found in both normal and malignant cell lines  |                |
| REFERENCE         |   | S40773   |                |
| #authors          |   | Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.   |                |
| #submission       |   | submitted to the EMBL Data Library, August 1990  |                |
| #accession        |   | S40773   |                |
|                   |   | ##molecule_type DNA  |                |
|                   |   | ##residues 1-393 ##label CHU   |                |
|                   |   | #cross-references EMBL:X54156; NID:q35213; PID:q35214  |                |

REFERENCE  
#authors S42669  
#journal Matlashewski, G.; Lamb, P.; Pim, D.; Peacock, J.; Crawford,  
L.; Benchimol, S.  
#title EMBO J. (1984) 3:3257-3262  
#accession Isolation and characterization of a human p53 cDNA clone:  
S42669 expression of the human p53 gene.  
#molecule-type mRNA  
#residues 101-393 ##label MK1  
#cross-references EMBL:X01405; NID:g35215; PID:g642241  
REFERENCE  
#authors Zakut-Houri, R.; Bienz-Tadmir, B.; Givol, D.; Oren, M.  
#journal EMBO J. (1985) 4:1251-1255  
#title Human p53 cellular tumor antigen: cDNA sequence and  
expression in COS cells.  
#cross-references MUID:85230577  
#accession A22837  
#molecule-type mRNA  
#residues 1-71, 'P', 73-393 ##label ZAK  
#cross-references EMBL:X02469; EMBL:M60950; NID:g35209; PID:g35210  
REFERENCE  
#authors Harlow, E.; Williamson, N.M.; Ralston, R.; Helfman, D.M.;  
Adams, T.E.  
#journal Mol. Cell. Biol. (1985) 5:1601-1610  
#title Molecular cloning and in vitro expression of a cDNA clone for  
human cellular tumor antigen p53.  
#accession A55060  
#molecule-type mRNA  
#residues 1-71, 'P', 73-272, 'H', 274-393 ##label HA3  
#cross-references GB:K03199; NID:g189478; PID:g189479  
#experimental-source clone p4-2, cell line A431  
REFERENCE  
#authors Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.;  
Arai, N.; Rotter, V.  
#journal Mol. Cell. Biol. (1986) 6:4650-4656  
#title Molecular basis for heterogeneity of the human p53 protein.  
#cross-references MUID:87089826  
#accession A25397  
#molecule-type mRNA  
#residues 1-78, 'T', 80-393 ##label HAR  
#cross-references EMBL:M14694; NID:g339613; PID:g339614  
#experimental-source clone p53-H-1, transformed hybridoma SV-80 cell  
line  
#accession B25397  
#molecule-type mRNA  
#residues 1-71, 'P', 73-78, 'T', 80-393 ##label HA2  
#cross-references EMBL:M14695; NID:g339615; PID:g339616  
#experimental-source clone p53-H-19, transformed hybridoma SV-80 cell  
line  
REFERENCE  
#authors Matlashewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider,  
J.; Crawford, L.V.  
#journal Mol. Cell. Biol. (1987) 7:961-963  
#title Primary structure polymorphism at amino acid residue 72 of  
human p53.  
#accession S42452  
#molecule-type DNA  
#residues 66-71, 'P', 73-79 ##label MK2  
#experimental-source clone lambda C113  
#note 72-Cys was also found, and appears to represent a  
polymorphism  
#accession S42453  
#molecule-type DNA  
#residues 66-79 ##label MAT  
#experimental-source clone J6K  
REFERENCE  
#authors Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.;  
Crook, T.  
#journal EMBO J. (1991) 10:2879-2887  
#title p53 is frequently mutated in Burkitt's lymphoma cell lines.  
#cross-references MUID:92007731  
#accession I38082  
#status translated from GB/EMBL/DBJ

#molecule-type mRNA  
#residues 1-189, 'L', ILSISEMKIECVMSIMVTEFLPDIVWCMPSRLRLALT',  
'VPSSTTTCVTPAPAA' ##label F01  
#cross-references EMBL:X60010; NID:g506432; PID:g506433  
#note deletion of a C nucleotide causes a frameshift at  
position 566  
#accession I38083  
#status translated from GB/EMBL/DBJ  
#molecule-type mRNA  
#residues 1-192, 'R', 194-393 ##label F02  
#cross-references EMBL:X60011; NID:g506434; PID:g506435  
#accession I38084  
#status translated from GB/EMBL/DBJ  
#molecule-type mRNA  
#residues 1-393 ##label F03  
#cross-references EMBL:X60012; NID:g506436; PID:g506437  
#accession I38085  
#status translated from GB/EMBL/DBJ  
#molecule-type mRNA  
#residues 1-245, 'T', 247-393 ##label F04  
#cross-references EMBL:X60013; NID:g506438; PID:g506439  
#accession I38086  
#status translated from GB/EMBL/DBJ  
#molecule-type mRNA  
#residues 1-236, 'T', 238-393 ##label F05  
#cross-references EMBL:X60014; NID:g506440; PID:g506441  
#accession I38087  
#status translated from GB/EMBL/DBJ  
#molecule-type mRNA  
#residues 1-247, 'Q', 249-393 ##label F06  
#cross-references EMBL:X60015; NID:g506442; PID:g506443  
#accession I38088  
#status translated from GB/EMBL/DBJ  
#molecule-type mRNA  
#residues 1-71, 'P', 73-237, 'Y', 239-393 ##label F07  
#cross-references EMBL:X60016; NID:g506444; PID:g506445  
#accession I38089  
#status translated from GB/EMBL/DBJ  
#molecule-type mRNA  
#residues 1-247, 'Q', 249-393 ##label F08  
#cross-references EMBL:X60017; NID:g506446; PID:g506447  
#accession I38090  
#status translated from GB/EMBL/DBJ  
#molecule-type mRNA  
#residues 1-71, 'P', 73-162, 'H', 164-393 ##label F09  
#cross-references EMBL:X60018; NID:g506448; PID:g506449  
#accession I38091  
#status translated from GB/EMBL/DBJ  
#molecule-type mRNA  
#residues 1-212, 'Q', 214-393 ##label F10  
#cross-references EMBL:X60019; NID:g506450; PID:g506451  
#accession I38092  
#status translated from GB/EMBL/DBJ  
#molecule-type mRNA  
#residues 1-253, 'D', 255-393 ##label F11  
#cross-references EMBL:X60020; NID:g506452; PID:g506453  
#note all sequences submitted to the EMBL/Genbank/DBJ  
databases June 1991  
REFERENCE  
#authors Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.  
#journal Nucleic Acids Res. (1991) 19:6977  
#title An Alu polymorphism intragenic to the TP53 gene.  
#cross-references MUID:92107726  
#accession I38093  
#status translated from GB/EMBL/DBJ  
#molecule-type DNA  
#residues 1-393 ##label RE2  
#cross-references EMBL:X54156; NID:g35213; PID:g35214  
REFERENCE  
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;  
Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.;  
Terada, M.  
#journal Cancer Res. (1991) 51:5800-5805

#title p53 gene mutations in gastric cancer metastases and in  
#cross-references EMBL:92034678  
#accession A44905  
##molecule\_type DNA  
##residues 246-247, 'W', 249-250 #label YAM  
##cross-references GB:S63157; NID:9237829; PID:9237830  
##note sequence extracted from NCBI backbone (NCBIN:63157,  
Note: remainder of annotations omitted.

Query Match 100.0%; Score 74; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 3,27e-02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPASWPL 93  
QY 1 APAPASWPL 10

RESULT 2  
ENTRY S06594 #type complete  
TITLE cellular tumor antigen p53 - green monkey  
ORGANISM #formal\_name Cercopithecus aethiops #common\_name green  
monkey, grivet

DATE 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change  
08-Sep-1997

ACCESSIONS  
REFERENCE S06594  
S06594

#authors Rigaudy, P.; Eckhart, W.  
#journal Nucleic Acids Res. (1989) 17:8375

#title Nucleotide sequence of a cDNA encoding the monkey cellular  
#cross-references NID:90045967

#accession S06594  
##molecule\_type mRNA

##residues 1-393 #label RIG  
##cross-references EMBL:X16384; NID:922795; PID:922796

CLASSIFICATION #superfamily cellular tumor antigen p53  
apoptosis; cell division control; DNA binding; homotrimer;  
nucleus; phosphoprotein; transcription regulation; tumor  
suppressor; zinc

FEATURE 176,179,236,242 #binding\_site zinc (Cys, His, Cys, Cys) #status  
predicted

392 #binding\_site phosphoryl-RNA (Ser) (covalent) #status  
predicted

SUMMARY #length 393 #molecular-weight 43696 #checksum 4263

Query Match 100.0%; Score 74; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 3,27e-02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPASWPL 93  
QY 1 APAPASWPL 10

RESULT 3  
ENTRY JC6193  
TITLE tumor suppressor p53 - rabbit  
ORGANISM #formal\_name Oryctolagus cuniculus #common\_name domestic  
rabbit

DATE 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change  
08-Sep-1997

ACCESSIONS  
REFERENCE JC6193  
JC6193

#authors Le Gos, F.; May, P.; Ronco, P.; de Fromental, C.C.  
#journal Gene (1997) 185:169-173

#title cDNA cloning and immunological characterization of rabbit  
#accession JC6193  
##molecule\_type mRNA

##residues 1-391 #label LEA  
##cross-references EMBL:X90592; NID:91532043; PID:9194962; PID:91532044

GENETICS  
#gene p53  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS tumor

SUMMARY #length 391 #molecular-weight 43435 #checksum 4367

Query Match 89.2%; Score 66; DB 2; Length 391;  
Best Local Similarity 90.0%; Pred. No. 5,50e-01;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 81 APAPASWPL 90  
QY 1 APAPASWPL 10

RESULT 4  
ENTRY S51648 #type complete  
TITLE cellular tumor antigen p53 - bovine  
ALTERNATE\_NAMES tumor-suppressor protein p53  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change  
08-Sep-1997

ACCESSIONS  
REFERENCE S51648  
S51648  
#authors Deguede, F.; Willems, L.; Burny, A.; Kettmann, R.  
#submission Submitted to the EMBL Data Library, September 1994  
#description Nucleotide sequence of the ovine p53 tumor-suppressor gene  
cDNA and its genomic organisation.

#accession S51648  
##molecule\_type mRNA  
##residues 1-386 #label DEQ  
##cross-references EMBL:X81704; NID:9602332; PID:9602333

CLASSIFICATION #superfamily cellular tumor antigen p53  
apoptosis; cell division control; DNA binding; homotrimer;  
phosphoprotein; transcription regulation; tumor suppressor;  
zinc

FEATURE 168,171,231,235 #binding\_site zinc (Cys, His, Cys, Cys) #status  
predicted

385 #binding\_site phosphoryl-RNA (Ser) (covalent) #status  
predicted

SUMMARY #length 386 #molecular-weight 43255 #checksum 7025

Query Match 86.5%; Score 64; DB 2; Length 386;  
Best Local Similarity 80.0%; Pred. No. 1,09e+00;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 76 TPAPASWPL 85  
QY 1 APAPASWPL 10

RESULT 5  
ENTRY S38824 #type complete  
TITLE cellular tumor antigen p53, alternative splice form - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change  
25-Oct-1996

ACCESSIONS  
REFERENCE S38822  
S38822  
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;  
Shohat, O.; Rotter, V.  
#journal Mol. Cell. Biol. (1986) 6:3232-3239

#title Immunologically distinct p53 molecules generated by  
alternative splicing.  
#accession S38824  
##molecule\_type mRNA  
##residues 1-381 #label ARA  
#reference S35478  
Han, K.A.; Kulcsz-Martin, M.F.

```

#journal      Nucleic Acids Res. (1992) 20:1979-1981
#title       Alternatively spliced p53 RNA in transformed and normal cells
#accession   S35478
#status      nucleic acid sequence not shown; translation not shown
#molecule_type mRNA
#residues    1-381
#label       HAN
#cross-references EMBL:M13874; NID:q200202; PID:q200203
#note        the nucleotide sequence was submitted to the EMBL Data
              Library, July 1988

COMMENT      This sequence, produced by alternative splicing of the tenth
              intron, lacks the carboxyl-terminal sequence necessary for
              covalent attachment of RNA. The function of this minor splice
              form is not known.

CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS       alternative splicing; phosphoprotein
FEATURES
1-44
16-26         #domain transcription activation #status predicted
99-289        #region conserved region I\
108-121       #domain DNA-binding core #status predicted #label DBC\
114-139       #region I1 loop\
160-192       #region conserved region II\
168-178       #region I2 loop\
231-252       #region conserved region III\
233-248       #region conserved region IV\
267-283       #region conserved region V\
313-319       #region nuclear location signal\
319-357       #region tetramer association\
7,9,12,18,23,37 #binding-site phosphate (Ser) (covalent) #status
              predicted\
173,176,235,239 #binding-site zinc (Cys, His, Cys, Cys) #status
              predicted\
312           #binding-site phosphate (Ser) (covalent) (by cdcc
              kinase) #status predicted
SUMMARY       #length 381 #molecular-weight 42498 #checksum 8703

Query Match      85.1%; Score 63; DB 2; Length 381;
Best Local Similarity 80.0%; Pred. No. 1.51e+00;
Matches          8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 81 APAPATPWL 90
OY 1 APAPASWPL 10

RESULT 6
ENTRY   DNMS53 #type complete
TITLE   cellular tumor antigen p53 - mouse
ALTERNATE_NAMES oncoprotein p53
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE     28-Aug-1985 #sequence_revision 04-Oct-1996 #text_change
          05-Sep-1997
ACCESSIONS A22733; S06336; A02684; S38822; S38823; I48703
REFERENCE  A22739
#authors   Blenz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.
#journal   EMBO J. (1984) 3:2179-2183
#cross-references MUID:85027173
#accession A22739
#molecule_type DNA
#residues  1-134, 'V', 136-390 #label BIE
REFERENCE  S06336
#authors   Chumakov, P.M.
#journal   Biochem. Khim. (1987) 13:1691-1694
#title     Primary structure of DNA complementary to murine oncoprotein
          p53 mRNA.
#cross-references MUID:88221682
#accession S06336
#molecule_type mRNA
#residues  1-134, 'V', 136-390 #label CHU
REFERENCE  A02684

```

```

#authors      Zakht-Houri, R.; Oren, M.; Blenz, B.; Lavie, V.; Hazum, S.;
#journal      Givol, D.
#title        Nature (1983) 306:594-597
#cross-references EMBL:1983; NID:95370; PID:95371
#accession    A02684
#molecule-type mRNA
#residues     1-159, 'H', 161-167, 'G', 169-233, 'T', 235-390 ##label ZAK
REFERENCE
#authors      Aral, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
#journal      Shohet, N.; Rotter, V.
#title        Mol. Cell. Biol. (1986) 6:3232-3239
#cross-references EMBL:M13872; NID:9200198; PID:9200199
#accession    S38822
#molecule-type mRNA
#status       preliminary
#residues     1-390 ##label ARA
#cross-references EMBL:M13872; NID:9200198; PID:9200199
#accession    S38823
#status       preliminary
#molecule-type mRNA
#residues     1-167, 'G', 169-233, 'T', 235-390 ##label AR2
#cross-references EMBL:M13873
REFERENCE
#authors      Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.
#journal      Nucleic Acids Res. (1984) 12:5609-5626
#title        Cloning and expression analysis of full length mouse cDNA
#cross-references EMBL:M13873
#cross-references MIM:84272240
#accession    I48703
#status       preliminary; translated from GB/EMBL/DBJ
#molecule-type mRNA
#residues     1-47, 'R', 49-78, 'W', 82-390 ##label RES
#cross-references EMBL:X00741; NID:953370; PID:953371
COMMENT
This DNA-binding protein plays an essential role in the regulation
of cell division, as it is required for the transition from phase
G0 to G1 of the cell cycle.
COMMENT
The tetramer association region may exhibit a beta-turn,
beta-sheet, beta-turn, alpha-helix motif.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS        apoptosis; cell division control; DNA binding; homotetramer;
                zincophorein; transcription regulation; tumor suppressor;
                zinc
FEATURE
1-44            #domain transcription activation #status predicted
                #label TRA\
16-26           #region conserved region I\
99-289          #domain DNA-binding core #status predicted #label DBC\
108-121         #region L1 loop\
114-139         #region conserved region II\
160-192         #region L2 loop\
168-178         #region conserved region III\
231-252         #region conserved region IV\
233-248         #region L3 loop\
267-283         #region conserved region V\
313-319         #region nuclear location signal\
319-357         #region tetramer association\
7,9,12,18,23,37 #binding_site phosphate (Ser) (covalent) #status
                predicted\
173,176,235,239 #binding_site zinc (Cys, His, Cys, Cys) #status
                predicted\
312             #binding_site phosphate (Ser) (covalent) (by cdcd
                kinase) #status predicted\
389             #binding_site phosphoryl-RNA (Ser) (covalent) #status
                predicted
SUMMARY
length 390 #molecular-weight 43458 #checksum 1260
Query Match      85.1%: Score 63; DB 1; length 390;
Best Local Similarity 80.0%: Pred. NO. 1,53e+00;
Matches          8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

Db 81 APAPATPMP 90  
|:|:|:|:|  
QY 1 APAPATPMP 10

RESULT 7  
ENTRY 7  
TITLE S02192 #type complete  
ALTERNATE\_NAMES cellular tumor antigen p53 - rat  
ORGANISM #gene p53 protein; nuclear oncoprotein p53  
#formal\_name Rattus norvegicus #common\_name Norway rat  
18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change  
08-Sep-1997  
ACCESSIONS S02192; S41149  
REFERENCE S02192  
#authors Soussi, T.; de Fromental, C.C.; Breugnot, C.; May, E.  
#journal Nucleic Acids Res. (1988) 16:11384  
#title Nucleotide sequence of a cDNA encoding the rat p53 nuclear  
oncoprotein.  
#cross-references MUID:89083585  
#accession S02192  
##molecule\_type mRNA  
##residues 1-391 ##label SOU  
#cross-references EMBL:X13058; NID:956828; PID:956829  
REFERENCE S41149  
#authors Hulla, J.E.; Schneider, R.P.  
#journal Nucleic Acids Res. (1993) 21:713-717  
#title Structure of the rat p53 tumor suppressor gene.  
#accession S41149  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
##molecule\_type DNA  
##residues 1-173 'W', 175-391 ##label HUT  
#cross-references EMBL:L07909  
#note the nucleotide sequence was submitted to the EMBL Data  
Library, December 1992

GENETICS 25/2: 32/3: 123/3: 185/1: 259/2: 305/1: 329/3: 365/2  
#introns  
CLASSIFICATION #superfamily cellular tumor antigen p53  
#apoptosis; cell division control; DNA binding; homotetramer;  
nucleus; phosphoprotein; transcription regulation; tumor  
suppressor; zinc  
KEYWORDS

FEATURE 174,177,236,240 #binding\_site zinc (Cys, His, Cys, Cys) #status  
14,177,236,240 predicted  
390 #binding\_site phosphoryl-RNA (Ser) (covalent) #status  
predicted  
SUMMARY #length 391 #molecular-weight 43451 #checksum 7105

Query Match 75.7%; Score 56; DB 2; Length 391;  
Best Local Similarity 70.0%; Pred. No. 1.54e+01;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 82 APAPATPMP 91  
|:|:|:|:|  
QY 1 APAPATPMP 10

RESULT 8  
ENTRY 8  
TITLE JC6176 #type complete  
ALTERNATE\_NAMES tumor suppressor protein p53 - Chinese hamster  
ORGANISM #formal\_name Crictetus griseus #common\_name Chinese hamster  
11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change  
08-Sep-1997  
ACCESSIONS JC6176  
REFERENCE JC6176  
#authors Lee, H.; Iarner, J.M.; Hamlin, J.L.  
#journal Gene (1997) 184:177-183  
#title Cloning and characterization of Chinese hamster p53 cDNA.  
#contents liver  
#accession JC6176  
#molecule\_type mRNA  
#residues 1-393 ##label LEE

##cross-references GB:U50395; NID:91842229; PID:91842230  
COMMENT This protein is a multimer. It plays the central role in a complex  
DNA damage-sensing network. It binds to replication factor and-  
TATA-binding protein, and affects DNA replication, transcription,  
and recombination by protein/protein interactions.

GENETICS p53  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS liver; tumor  
SUMMARY #length 393 #molecular-weight 43362 #checksum 4043

Query Match 75.7%; Score 56; DB 2; Length 393;  
Best Local Similarity 70.0%; Pred. No. 1.54e+01;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 84 ASAPATPMP 93  
|:|:|:|:|  
QY 1 APAPATPMP 10

RESULT 9  
ENTRY 9  
TITLE JH0633 #type complete  
ALTERNATE\_NAMES cellular tumor antigen p53 - golden hamster  
ORGANISM #formal\_name Mesocricetus auratus #common\_name golden hamster  
17-Aug-1992 #sequence\_revision 17-Aug-1992 #text\_change  
08-Sep-1997  
ACCESSIONS JH0633  
REFERENCE JH0633  
#authors Legros, Y.; Mcintyre, P.; Soussi, T.  
#journal Gene (1992) 112:247-250  
#title The cDNA cloning and immunological characterization of  
hamster p53.  
#cross-references MUID:92210007  
#accession JH0633  
##molecule\_type mRNA  
##residues 1-396 ##label LEG  
#cross-references GB:M75144; NID:9191414; PID:9191415  
#experimental\_source kidney, strain MPI

GENETICS p53  
CLASSIFICATION #superfamily cellular tumor antigen p53  
#apoptosis; cell division control; DNA binding; homotetramer;  
nucleus; phosphoprotein; transcription regulation; tumor  
suppressor; zinc  
KEYWORDS

FEATURE 179,182,241,245 #binding\_site zinc (Cys, His, Cys, Cys) #status  
179,182,241,245 predicted  
395 #binding\_site phosphoryl-RNA (Ser) (covalent) #status  
predicted  
SUMMARY #length 396 #molecular-weight 43631 #checksum 6617

Query Match 75.7%; Score 56; DB 2; Length 396;  
Best Local Similarity 70.0%; Pred. No. 1.54e+01;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 87 ASAPATPMP 96  
|:|:|:|:|  
QY 1 APAPATPMP 10

RESULT 10  
ENTRY 10  
TITLE B43776 #type complete  
ALTERNATE\_NAMES drebrin E1 - chicken  
ORGANISM #formal\_name Gallus gallus #common\_name chicken  
01-Dec-1992 #sequence\_revision 30-Jan-1993 #text\_change  
30-Sep-1993  
ACCESSIONS B43776  
REFERENCE B43776  
#authors Kojima, N.; Kato, Y.; Shiro, T.; Obata, K.  
#journal Brain Res. Mol. Brain Res. (1988) 4:207-215  
#title Nucleotide sequences of two embryonic drebrins,  
developmentally regulated brain proteins, and developmental

change in their mRNAs.

##accession B43776  
##status preliminary  
##molecule\_type mRNA  
##residues 1-564 ##label KOJ  
SUMMARY #length 564 #molecular\_weight 62296 #checksum 8914

Query Match 75.7%; Score 56; DB 2; Length 564;  
Best Local Similarity 60.0%; Pred. No. 1.54e+01;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db: 408 APAATSWPL 417  
OY 1 APAPASWPL 10

RESULT 11  
ENTRY I51213 #type fragment  
TITLE drebrln - chicken (fragment)  
ORGANISM #formal\_name Gallus gallus #common\_name chicken  
DATE 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 07-Nov-1997

ACCESSIONS I51213  
REFERENCE I51212  
#authors Kojima, N.; Shiroo, T.; Obata, K.  
#journal Brain Res. Mol. Brain Res. (1993) 19:101-114  
#title Molecular cloning of a developmentally regulated brain protein, chicken drebrln A and its expression by alternative splicing of the drebrln gene.

#cross-references M01D:93368392

#accession I51213  
##status preliminary; translated from GB/EMBL/DBJ  
##molecule\_type DNA  
##residues 1-593 ##label KOJ  
#cross-references GB:S5296; NID:9410604; PID:9410605

GENETICS #introns 26/3; 51/3; 100/3; 126/3; 177/2; 198/3; 218/3; 257/1; 303/1;  
SUMMARY #length 593 #checksum 1479

Query Match 75.7%; Score 56; DB 2; Length 593;  
Best Local Similarity 60.0%; Pred. No. 1.54e+01;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db: 438 APAATSWPL 447  
OY 1 APAPASWPL 10

RESULT 12  
ENTRY A43776 #type complete  
TITLE drebrln E2 - chicken  
ORGANISM #formal\_name Gallus gallus #common\_name chicken  
DATE 01-Dec-1992 #sequence\_revision 30-Jan-1993 #text\_change 06-Dec-1996

ACCESSIONS A43776; I50221  
REFERENCE A43776  
#authors Kojima, N.; Kato, Y.; Shiroo, T.; Obata, K.  
#journal Brain Res. Mol. Brain Res. (1988) 4:207-215  
#title Nucleotide sequences of two embryonic drebrlins, developmentally regulated brain proteins, and developmental change in their mRNAs.

#accession A43776  
##molecule\_type mRNA  
##residues 1-607 ##label KOJ  
#cross-references GB:M36961; NID:9211725; PID:9211726  
SUMMARY #length 607 #molecular\_weight 66685 #checksum 2901

Query Match 75.7%; Score 56; DB 2; Length 607;  
Best Local Similarity 60.0%; Pred. No. 1.54e+01;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db: 451 APAATSWPL 460

OY 1 APAPASWPL 10

RESULT 13  
ENTRY C69187 #type complete  
TITLE conserved hypothetical protein MTH656 - Methanobacterium  
ORGANISM #formal\_name Methanobacterium (strain Delta H)  
DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Dec-1997

ACCESSIONS C69187  
REFERENCE A69000  
#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;

Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakey, D.;  
Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;  
Lumm, W.; Potlter, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;  
Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;  
McDougal, S.; Shimer, G.; Goyal, A.; Pietrovski, S.;  
Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

#journal J. Bacteriol. (1997) 179:7135-7155  
#title Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.

#cross-references M01D:98037514  
#accession C69187  
##status preliminary; nucleic acid sequence not shown; translation not shown

#molecule\_type DNA  
##residues 1-837 ##label MTH  
#cross-references GB:AE000666  
#experimental\_source strain Delta H

GENETICS #gene MTH656  
SUMMARY #start\_codon TTG #length 837 #molecular\_weight 93978 #checksum 6898

Query Match 75.7%; Score 56; DB 2; Length 837;  
Best Local Similarity 75.0%; Pred. No. 1.54e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db: 613 SPDSWPL 620  
OY 3 APAPSWPL 10

RESULT 14  
ENTRY TYF7L #type complete  
TITLE kinase-related protein sevenless - fruit fly (Drosophila melanogaster)  
CONTAINS protein-tyrosine kinase (EC 2.7.1.112)  
ORGANISM #formal\_name Drosophila melanogaster  
DATE 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 05-Sep-1997

ACCESSIONS A28912; A60107; A28827  
REFERENCE A28912  
#authors Basler, K.; Hafen, E.  
#journal Cell (1988) 54:299-311  
#title Control of photoreceptor cell fate by the sevenless protein requires a functional tyrosine kinase domain.

#cross-references M01D:88282538  
#accession A28912  
##molecule\_type DNA  
##residues 1-254 ##label BAS  
#cross-references GB:J03158; NID:9158418; PID:9158419

REFERENCE A60107  
#authors Hafen, E.; Basler, K.; Edstroem, J.E.; Rubin, G.M.  
#journal Science (1987) 236:55-63  
#title Sevenless, a cell-specific homeotic gene of Drosophila, encodes a putative transmembrane receptor with a tyrosine kinase domain.



#cross-references MUID:87177965  
 #accession A60107  
 #status nucleic acid sequence not shown  
 #molecule\_type DNA  
 #residues 2062-2554 ##label HAF

REFERENCE  
 #authors A28827  
 #journal Bowtell, D.D.L.; Simon, M.A.; Rubin, G.M.  
 #title Genes Dev. (1988) 2:620-634  
 Nucleotide sequence and structure of the sevenless gene of Drosophila melanogaster.

#cross-references MUID:88329706  
 #accession A28827  
 #molecule\_type DNA; mRNA  
 #residues 'RSSAS', 1-391, 'V', 393-1822, 'Q', 1824-2270, 'C', 2272-2554  
 ##label BOW

#cross-references GB:X13665  
 #note 392-Met, 1000-Leu, 1364-Val, 1668-Val, 1703-His,  
 1730-Lys, 1731-Glu, 1741-Met, and 2271-Arg were also  
 found

COMMENT The sevenless gene controls the development of the R7 class of photoreceptor cells.

GENETICS  
 #gene sev  
 #cross-references FlyBase:FBgn0003366  
 #map\_position X10A1-A2  
 #introns 227/3: 252/3: 274/3: 406/3: 486/3: 596/2: 2150/1: 2242/3;  
 2307/3: 2456/2

CLASSIFICATION  
 #superfamily sevenless; fibronectin type III repeat homology;  
 LDL receptor YMTD-containing repeat homology; protein  
 kinase homology  
 ATP: autophosphorylation; glycoprotein; phosphoprotein;  
 phosphotransferase; photoreceptor; receptor; transmembrane  
 protein; tyrosine-specific protein kinase

KEYWORDS

FEATURE  
 102-122 #domain transmembrane #status predicted #label TMN1\  
 437-528 #domain fibronectin type III repeat homology #label FN3\  
 2124-2147 #domain transmembrane #status predicted #label TMN2\  
 2207-2488 #domain protein kinase homology #label KIN\  
 2245-2223 #region protein kinase ATP-binding motif  
 129,481,505,617  
 647,966,1228,1313,  
 1353,1550,1557,  
 1639,1725,1756,  
 1804,1889,1947,  
 2073

2242 #binding-site, carbohydrate (Asn) (covalent) #status  
 predicted  
 #active-site Lys #status predicted  
 #length 2554 #molecular-weight 287108 #checksum 3353

SUMMARY  
 #length 2554 #molecular-weight 287108 #checksum 3353

Query Match 75.7%; Score 56; DB 1; Length 2554;  
 Best Local Similarity 60.0%; Pred. No. 1.54e+01;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 793 AASPASWPL 802  
 1:::|||||  
 QY 1 APAPASWPL 10

RESULT 15  
 ENTRY F39690 #type fragment  
 TITLE neural cell adhesion molecule, heart (splice form +, -, +, +) -  
 ORGANISM rat (fragment)  
 #formal\_name Rattus norvegicus #common\_name Norway rat  
 #journal 24-Jan-1992 #sequence\_revision 24-Jan-1992 #text\_change  
 07-Feb-1997

ACCESSIONS  
 F39690  
 REFERENCE A39690  
 #authors Reyes, A.A.; Small, S.J.; Akesson, R.  
 #journal Mol. Cell. Biol. (1991) 11:1654-1661  
 #title At least 27 alternatively spliced forms of the neural cell  
 adhesion molecule mRNA are expressed during rat heart  
 development.  
 #cross-references MUID:91141516

#accession F39690  
 #status preliminary; nucleic acid sequence not shown; not  
 compared with conceptual translation  
 #molecule\_type mRNA  
 #residues 1-27 ##label REY  
 #cross-references GB:M63970

CLASSIFICATION  
 #superfamily neural cell adhesion molecule; fibronectin type  
 III repeat homology; immunoglobulin homology  
 cell adhesion; heart

KEYWORDS  
 SUMMARY #length 27 #checksum 9769

Query Match 74.3%; Score 55; DB 2; Length 27;  
 Best Local Similarity 55.6%; Pred. No. 2.11e+01;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 5 SPPPTWPL 13  
 1:::|||||  
 QY 2 PAPASWPL 10

Search completed: Fri Sep 11 13:29:18 1998  
 Job time : 19 secs.



(TM)

Distribution rights by Oxford Molecular Ltd

Run on: Fri Sep 11 13:29:36 1998; MasPar time 2.08 Seconds

120.861 Million cell updates/sec

1 APAPAPSWPL 10

PAM 15C

69111 seqs, 25083644 residues

Minimum Match 0%  
Listing first 45 summaries

swiss-prot35  
1:swiss1

Mean 23.678; Variance 40.632; scale 0.583

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query | Length | DB | ID         | Description             | Pred. No. |
|------------|-------|-------|--------|----|------------|-------------------------|-----------|
| 1          | 74    | 100.0 | 393    | 1  | P53_CGAAE  | CELLULAR TUMOR ANTIGEN  | 4.11e-03  |
| 2          | 74    | 100.0 | 393    | 1  | P53_HUMAN  | CELLULAR TUMOR ANTIGEN  | 4.11e-03  |
| 3          | 71    | 95.9  | 276    | 1  | P53_CANFA  | CELLULAR TUMOR ANTIGEN  | 1.41e-02  |
| 4          | 66    | 89.2  | 280    | 1  | P53_HORSE  | CELLULAR TUMOR ANTIGEN  | 1.05e-01  |
| 5          | 66    | 89.2  | 391    | 1  | P53_RABIT  | CELLULAR TUMOR ANTIGEN  | 1.05e-01  |
| 6          | 64    | 86.5  | 386    | 1  | P53_BOVIN  | CELLULAR TUMOR ANTIGEN  | 2.31e-01  |
| 7          | 63    | 85.1  | 390    | 1  | P53_MOUSE  | CELLULAR TUMOR ANTIGEN  | 3.41e-01  |
| 8          | 60    | 81.1  | 314    | 1  | P53_SPEBE  | CELLULAR TUMOR ANTIGEN  | 1.08e+00  |
| 9          | 60    | 81.1  | 386    | 1  | P53_FELCA  | CELLULAR TUMOR ANTIGEN  | 1.08e+00  |
| 10         | 58    | 78.4  | 895    | 1  | GND5_RAT   | GUININE NUCLEOTIDE DIS  | 2.27e+00  |
| 11         | 56    | 75.7  | 391    | 1  | P53_RAT    | CELLULAR TUMOR ANTIGEN  | 4.75e+00  |
| 12         | 56    | 75.7  | 393    | 1  | P53_CRIGR  | CELLULAR TUMOR ANTIGEN  | 4.75e+00  |
| 13         | 56    | 75.7  | 396    | 1  | P53_MESAU  | CELLULAR TUMOR ANTIGEN  | 4.75e+00  |
| 14         | 56    | 75.7  | 652    | 1  | DREB_CHICK | DREBRINS A, E1 AND E2.  | 4.75e+00  |
| 15         | 56    | 75.7  | 2554   | 1  | 7ZES_DROME | SEVENTEEN PROTEIN (EC   | 4.75e+00  |
| 16         | 54    | 73.0  | 212    | 1  | E11_HUMAN  | ENOETHHELIN-1 PRECURSOR | 9.76e+00  |
| 17         | 54    | 73.0  | 268    | 1  | CDX1_MOUSE | HOMEOBOX PROTEIN CXX-1  | 9.76e+00  |
| 18         | 54    | 73.0  | 382    | 1  | P53_SHEEP  | CELLULAR TUMOR ANTIGEN  | 9.76e+00  |
| 19         | 54    | 73.0  | 440    | 1  | CDX2_MOUSE | PUTATIVE TACRYKLININ RE | 9.76e+00  |
| 20         | 53    | 71.6  | 311    | 1  | TKXR_HUMAN | HOMEOBOX PROTEIN CXX-2  | 1.39e+01  |
| 21         | 53    | 71.6  | 311    | 1  | CDX2_HUMAN | HOMEOBOX PROTEIN CXX-2  | 1.39e+01  |
| 22         | 53    | 71.6  | 670    | 1  | DSH_HUMAN  | SEGMENT POLARITY PROTE  | 1.39e+01  |
| 23         | 51    | 68.9  | 257    | 1  | YOS6_MITCU | HYPOTHELTICAL 27.6 KD P | 2.80e+01  |

|    |                        |          |
|----|------------------------|----------|
| 24 | SPB, STRI              | 2.80e+01 |
| 25 | VI2, HPV6              | 2.80e+01 |
| 26 | MINOR CASPID PROTEIN L | 2.80e+01 |
| 27 | GLYCOPROTEIN E PRECURS | 2.80e+01 |
| 28 | KITR_HUMAN             | 2.80e+01 |
| 29 | LEUKOCYTE TYROSINE KIN | 2.80e+01 |
| 30 | DAB, DROM              | 2.80e+01 |
| 31 | HYPOHETICAL 30.6 KD P  | 3.94e+00 |
| 32 | GLYG_MYCTO             | 3.94e+00 |
| 33 | GLYG_HUMAN             | 3.94e+00 |
| 34 | GLYG_RAIT              | 3.94e+00 |
| 35 | GLYG_MYCO              | 3.94e+00 |
| 36 | TRANSPOSASE FOR INSERT | 3.94e+00 |
| 37 | POLR, EPAP             | 3.94e+00 |
| 38 | RNA REPLICASE POLYPROT | 3.94e+00 |
| 39 | LACTAM UTILIZATION PRO | 5.52e+01 |
| 40 | OVARY MATURING PARSII  | 5.52e+01 |
| 41 | TRAA, ECOLI            | 5.52e+01 |
| 42 | TRAA, ECOLI            | 5.52e+01 |
| 43 | TERAHYDROMETHANOPERI   | 5.52e+01 |
| 44 | HYPOHETICAL 41.8 KD P  | 5.52e+01 |
| 45 | Y4KF_RHISN             | 5.52e+01 |
| 46 | HYPOHETICAL 41.8 KD P  | 5.52e+01 |
| 47 | CP80_BEST              | 5.52e+01 |
| 48 | BEEDBANUNINE SYNTHASE  | 5.52e+01 |
| 49 | GND5_MOUSE             | 5.52e+01 |
| 50 | GUANINE NUCLEOTIDE DIS | 5.52e+01 |
| 51 | KITR_MOUSE             | 5.52e+01 |
| 52 | LEUKOCYTE TYROSINE KIN | 5.52e+01 |
| 53 | IRBP_HUMAN             | 5.52e+01 |
| 54 | INTERPHOSRECEPTOR RET  | 7.70e+01 |
| 55 | RENIN PRECURSOR, RENAL | 7.70e+01 |
| 56 | DMP2_HUMAN             | 7.70e+01 |
| 57 | DYSTROPHIN-RELATED PRO | 7.70e+01 |

## ALIGNMENTS

| ID | RESULT   | 1                                 | STANDARD: | PRT:                                     | 393 AA. |
|----|--|-----------------------------------|-----------|--|---------|
| AC | P53  | CERAE                             |           |  |         |
| AC | P13481:  |                                   |           |  |         |
| DT | 01-JAN-1990  | (REL. 13, CREATED)                |           |  |         |
| DT | 01-JAN-1990  | (REL. 13, LAST SEQUENCE UPDATE)   |           |  |         |
| DT | 01-NOV-1997  | (REL. 35, LAST ANNOTATION UPDATE) |           |  |         |
| DE | CELLULAR TUMOR ANTIGEN P53.  |                                   |           |  |         |
| GN | TP53.  |                                   |           |  |         |
| OS | CERCOPITHECUS AETHIOPS (GREEN MONKEY) (GRIVET).                      |                                   |           |  |         |
| OC | EUKARYOTA; METAZOA; CHORALTA; VERTEBRATA; TETRAPODA; MAMMALIA;       |                                   |           |  |         |
| OC | EUHAEIA; PRIMATES.   |                                   |           |  |         |
| RN | [1]  |                                   |           |  |         |
| RP | SEQUENCE FROM N.A.   |                                   |           |  |         |
| RC | TISSUE=LIVER;  |                                   |           |  |         |
| RX | MEDLINE: 90045967.   |                                   |           |  |         |
| RA | RIGAUDY P., ECKHARDT W.;   |                                   |           |  |         |
| RL | NUCLEIC ACIDS RES. 17:8375-8375(1989).                               |                                   |           |  |         |
| CC | -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES |                                   |           |  |         |
| CC | GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL            |                                   |           |  |         |
| CC | CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN      |                                   |           |  |         |
| CC | TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A         |                                   |           |  |         |
| CC | TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION   |                                   |           |  |         |
| CC | BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF      |                                   |           |  |         |
| CC | THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.     |                                   |           |  |         |
| CC | APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF    |                                   |           |  |         |
| CC | BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2            |                                   |           |  |         |
| CC | EXPRESSION.  |                                   |           |  |         |
| CC | -1- SUBCELLULAR LOCATION: NUCLEAR.                                   |                                   |           |  |         |
| CC | -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY     |                                   |           |  |         |
| CC | OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED       |                                   |           |  |         |
| CC | IN MANY TYPES OF CANCER.   |                                   |           |  |         |
| CC | -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.           |                                   |           |  |         |
| DR | EMBL: X16384; G22796; .  |                                   |           |  |         |
| DR | PIR: S06594; S06594.   |                                   |           |  |         |
| DR | HSSP: P04637; IOLG.  |                                   |           |  |         |
| DR | PROSITE: PS00348; P53: 1.  |                                   |           |  |         |
| DR | ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;     |                                   |           |  |         |
| KW | NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.                         |                                   |           |  |         |
| FT | DOMAIN   | 1                                 | 68        | ASP/GLU-RICH (ACIDIC).                   |         |
| FT | DOMAIN   | 81                                | 150       | HYDROPHOBIC.                             |         |
| FT | DOMAIN   | 319                               | 393       | HIGHLY BASIC AND MAY BE INVOLVED IN      |         |
| FT | DOMAIN   |                                   |           | INTERACTION WITH DNA.                    |         |
| FT | DOMAIN   |                                   |           | NUCLEAR LOCALIZATION SIGNAL (POTENTIAL). |         |
| FT | MOD_RES  | 311                               | 323       | PHOSPHORYLATION (BY SIMILARITY).         |         |
| FT | MOD_RES  | 392                               | 392       |  |         |
| SO | SEQUENCE   | 393 AA:                           | 43696 MW; | BEE7DC62 CRC32:                          |         |

Query Match 100.0%; Score 74; DB 1; Length 393;

Best Local Similarity 100.0%; Pred. No. 4,11e-03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

db 84 APAPSPWPL 93  
1111111111  
1 APAPSPWPL 10

RESULT 2  
ID P53\_HUMAN STANDARD: PRT: 393 AA.  
AC P04637;  
DT 13-AUG-1987 (REL. 05, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
GN TP53.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 8523057.  
RA ZAKUT-HOURI R., BIENZ-TADMOR B., GIVOL D., OREN M.;  
ENBO J. 4:1251-1255(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87064416.  
RA LAMB P., CRAWFORD L.;  
MOL. CELL. BIOL. 6:1379-1385(1986).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85267676.  
RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;  
MOL. CELL. BIOL. 5:1601-1610(1985).  
RN [4]  
RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
RX MEDLINE: 87089826.  
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
ROTTER V.;  
MOL. CELL. BIOL. 6:4650-4656(1986).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 89108008.  
RA BUCHANAN V.L., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P.,  
GEOORGIEV G.P.;  
GENE 70:245-252(1988).  
RN [6]  
RP SEQUENCE OF 101-393 FROM N.A.  
RX MEDLINE: 85126934.  
RA MATLASHENSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
BENCHIMOL S.;  
EMBO J. 3:3257-3262(1984).  
RN [7]  
RP NUCLEAR LOCALIZATION SIGNAL.  
RX MEDLINE: 90191730.  
RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
ONCOGENE 5:423-426(1990).  
RN [8]  
RP STRUCTURE BY NMR OF 319-360.  
RX MEDLINE: 94294808.  
RA GLORE G.M., OMICHENSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
APPELLA E., GRONENBORN A.M.;  
SCIENCE 265:386-391(1994).  
RN [9]  
RP STRUCTURE BY NMR OF 325-355.  
RX MEDLINE: 95292092.  
RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
NAT. STRUCT. BIOL. 1:877-890(1994).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
RX MEDLINE: 94294806.  
RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
SCIENCE 265:346-355(1994).

RN [11]  
RP REVIEW.  
RX MEDLINE: 94090335.  
RA HARRIS C.C.;  
SCIENCE 262:1980-1981(1993).  
RN [12]  
RP REVIEW ON VARIANTS.  
RX MEDLINE: 91289156.  
RA HOULSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
SCIENCE 253:49-53(1991).  
RN [13]  
RP REVIEW ON VARIANTS.  
RX MEDLINE: 96271983.  
RA DE VRIES E.M.G., RICE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
LIAO D., SOUSSI T., KOVACH J.S., SOMMER S.S.;  
HUM. MUTAT. 7:202-213(1996).  
RN [14]  
RP VARIANT ARG-72.  
RX MEDLINE: 91153807.  
RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
HUM. GENET. 86:369-370(1991).  
RN [15]  
RP VARIANT LFS THR-133.  
RX MEDLINE: 92034774.  
RA LAU J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
CANCER RES. 51:6385-6387(1991).  
RN [16]  
RP VARIANTS LFS CYS-245, TRP-248, PRO-252 AND LYS-258.  
RX MEDLINE: 91057657.  
RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
KIM D.H., KASSEL J., GRAY M.A., BISCHOFF F.Z., TAINSKY M.A.,  
FRIEND S.H.;  
SCIENCE 250:1233-1238(1990).  
RN [17]  
RP VARIANT LFS ASP-245.  
RX MEDLINE: 91080929.  
RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
NATURE 348:747-749(1990).  
RN [18]  
RP VARIANT LFS LEU-272.  
RX MEDLINE: 92147883.  
RA FELIX C.A., NAV M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
KNUTSEN T., MINNA J.D.;  
J. CLIN. INVEST. 89:640-647(1992).  
RN [19]  
RP VARIANTS LFS HIS-273 AND VAL-325.  
RX MEDLINE: 92228023.  
RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
GEHARAD M.C., ANDERSEN T.I., BORESEN A.-L., LI F.P., GARBER J.,  
STRONG L.C.;  
NEW ENGL. J. MED. 326:1309-1315(1992).  
RN [20]  
RP VARIANTS BREAST TUMORS GLN-132, SER-249, LYS-280 AND LYS-285.  
RX MEDLINE: 90295284.  
RA BARTEK J., IGGO R., GANNON J., LANE D.P.;  
ONCOGENE 5:893-899(1990).  
RN [21]  
RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
RX MEDLINE: 91017544.  
RA RODRIGUES N.R., ROMAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
GANNON J.V., LANE D.P.;  
PROC. NATL. ACAD. SCI. U.S.A. 87:7555-7559(1990).  
RN [22]  
RP VARIANTS COLORECTAL CANCER MUTATIONS.  
RX MEDLINE: 91282784.  
RA ISHIOKA C., SATO T., GAMOH M., SUZUKI T., SHIBATA H., KANAMARU R.,  
MAKUI A., YAMAZAKI T.;  
BIOCHEM. BIOPHYS. RES. COMMUN. 177:901-906(1991).  
RN [23]  
RP VARIANTS OESOPHAGUS TUMORS L-152; A-155; H-175; F-176 AND H-273.  
RX MEDLINE: 91330175.  
RA CASSON A.G., MUKHOPADHYAY T., CLEARY K.R., RO J.Y., LEVIN B.,

RA ROTH J. A.;  
RN CANCER RES. 51:4495-4499(1991).  
RP  
RX  
RN [24]  
RP VARIANTS IN HEPATOCELLULAR CARCINOMAS IN CHINA.  
RX MEDLINE; 91187113.  
RA HSU I. C., METCALF R. A., SUN T., WELSH J. A., WANG N. J., HARRIS C. C.;  
RL NATURE 350:427-428(1991).  
RN [25]  
RP VARIANTS IN HEPATOCELLULAR CARCINOMAS IN SOUTH AFRICA.  
RX MEDLINE; 91187114.  
RA BRESSAC B., KEM M., WANDS J., OZTURK M.;  
RN NATURE 350:429-431(1991).  
RP [26]  
RX VARIANTS IN ANOGENITAL CARCINOMAS.  
RN MEDLINE; 93010989.  
RA CROOK T., VOUSDEN K. H.;  
RN EMO J. 11:3935-3940(1992).  
RP [27]  
RX VARIANT WITH DUPLICATION OF PRO-177-HIS-178.  
RN MEDLINE; 93265016.  
RA BHATIA K., GUTTEREZ M. I., MAGRATH I. T.;  
RN HUM. MOL. GENET. 1:207-208(1992).  
RP [28]  
RX VARIANTS IN BURKITT'S LYMPHOMAS.  
RN MEDLINE; 93064692.  
RA DUTHU A., DEBIRE B., ROMANO J. W., EHRHART J. C., FISCELLA M., MAY E.,  
RX APPELLA E., MAY P.;  
RN ONCOGENE 7:2161-2167(1992).  
RP [29]  
RX VARIANT IN NASOPHARYNGEAL CARCINOMA THR-280.  
RN MEDLINE; 92335329.  
RA SUN Y. Y., HEGAMER G., HENG Y. J., HILDESHEIM A., CHEN J. Y., CAO Y.,  
RX YAO C. T., COLEBURN N. H.;  
RN PROC. NATL. ACAD. SCI. U.S.A. 89:6516-6520(1992).  
RP [30]  
RX VARIANTS IN COLON TUMORS.  
RN MEDLINE; 93330562.  
RA HAMELIN R., JEGO N., LAURENT-PUIG P., VIDAUD M., THOMAS G.;  
RN ONCOGENE 8:2213-2220(1993).  
RP [31]  
RX CHARACTERIZATION OF VARIANT ALA-143.  
RN MEDLINE; 94283378.  
RA ZHANG W., GUO X. Y., HU G. Y., LIU W. B., SHAY J. W., DEISENROTH A. B.;  
RN EMO J. 13:2555-2544(1994).  
RP [32]  
RX VARIANTS LFS HIS-175; ARG-193; GIN-248; CYS-273 AND TYR-275.  
RN MEDLINE; 95193787.  
RA FERBOURG T., BARBIER N., YAN Y. X., GABER J. E., DREXFUS M.,  
RX FRAUMERI J. F. JR., LI F. P., FIEND S. H.;  
RN AM. J. HUM. GENET. 56:608-615(1995).  
RP [33]  
RX VARIANT LFS HIS-175.  
RN MEDLINE; 96423319.  
RA VARLEY J. M., MCGOWN G., THORNCROFT M., TRICKER K. J., TEARE M. D.,  
RX SANTIABNE-KOREF M. F., HOUSTON R. S., MARTIN J., BIRCH J. M.,  
RA EVANS D. G. R.;  
RN J. MED. GENET. 32:942-945(1995).  
RP [34]  
RX VARIANTS BA PHE-176; SER-245; TRP-248; TRP-282 AND GIN-286.  
RN MEDLINE; 96233927.  
RA ADREDET M. P., ROBASZKIEWICZ M., MERCIER B., NOUSBAUM J. B.,  
RX HARDY E., BALI J. P., VOLANT A., LOZACH P., GOUEROU H., FEREC C.;  
RN HUM. MUTAT. 7:109-113(1996).

```

Query Match      100.0%;      Score 74; DB 1; Length 39;
Best Local Similarity 100.0%;      Pred No. 4.11e-03;
Matches      10; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Db      84      AAPAPSPML 93
      1      | | | | | | | | | |
Qy      1      AAPAPSPML 10

```

| RESULT   | ID       | 3  | STANDARD; | PRT; | 276 AA.                |
|----------|----------|--|-----------|------|------------------------|
| AC       | AD       | P53 CANFA<br>Q29537;   |           |      |                        |
| DT       | DT       | 01-NOV-1997 (REL. 35, CREATED)                                       |           |      |                        |
| DT       | DT       | 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)                          |           |      |                        |
| DT       | DT       | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)                        |           |      |                        |
| DE       | DE       | CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).                               |           |      |                        |
| GN       | GN       | TP53.  |           |      |                        |
| OS       | OS       | CANIS FAMILIARIS (DOG).  |           |      |                        |
| OC       | OC       | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;       |           |      |                        |
| OC       | OC       | EUTHERIA; CARNIVORA.   |           |      |                        |
| RN       | RN       | [1]  |           |      |                        |
| RP       | RP       | SEQUENCE FROM N.A.   |           |      |                        |
| RC       | RC       | STRAIN-BEAGLE;   |           |      |                        |
| RX       | RX       | MEDLINE: 95323915.   |           |      |                        |
| RA       | RA       | KRAEGL S.A., PAZI K.A., MADEWELL B.R.;                               |           |      |                        |
| RL       | RL       | CANCER LETT. 92:181-186(1995).                                       |           |      |                        |
| CC       | CC       | -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES |           |      |                        |
| CC       | CC       | GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL            |           |      |                        |
| CC       | CC       | CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN      |           |      |                        |
| CC       | CC       | TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A         |           |      |                        |
| CC       | CC       | TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION   |           |      |                        |
| CC       | CC       | BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF      |           |      |                        |
| CC       | CC       | THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.     |           |      |                        |
| CC       | CC       | APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF    |           |      |                        |
| CC       | CC       | BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2            |           |      |                        |
| CC       | CC       | EXPRESSION.  |           |      |                        |
| CC       | CC       | -1- SUBCELLULAR LOCATION: NUCLEAR                                    |           |      |                        |
| CC       | CC       | -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY     |           |      |                        |
| CC       | CC       | OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED       |           |      |                        |
| CC       | CC       | IN MANY TYPES OF CANCER.   |           |      |                        |
| CC       | CC       | -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.           |           |      |                        |
| DR       | DR       | EMBL: S77819: G1000577; -.   |           |      |                        |
| DR       | DR       | PROSITE: PS00348; P53: 1.  |           |      |                        |
| DR       | DR       | ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;     |           |      |                        |
| KW       | KW       | NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.                         |           |      |                        |
| FT       | FT       | NON_TER  | 1         | 1    |                        |
| FT       | FT       | NON_TER  | <1        | 35   | ASP/GLU-RICH (ACIDIC). |
| FT       | FT       | NON_TER  | 276       | 276  |                        |
| SEQUENCE | SEQUENCE | 276 AA: 30466 MW: 8697AE44 CRC32:                                    |           |      |                        |

```

Query Match          95.9%: Score 71; DB 1; Length 276;
Best Local Similarity 90.0%: Pred. No. 1,41e-02;
Matches          9; Conservative          1; Mismatches          0; Indels          0; Gaps          0;

Db          47 APGAPSWPL 56
           |||iiiiii
QY          1 APAPPSWPL 10

RESULT      4
ID          P53_HORSE          STANDARD;          PRT;          280 AA.
AC          P79892: 029481:
DT          01-NOV-1997 (REL. 35, CREATED)
DT          01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT          01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE          CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN          TP53 OR P53.
OS          EODUS CABALLUS (HORSE).
OC          EOKAROTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:
OC          EUETHERIA: PERISSODACTYLA.
RN          [1]
RP          SEQUENCE OF 1-263 FROM N.A.
RC          TISSUE=SPLEEN;
RX          MEDLINE: 97070350.
RA          PAZZI K.A., KRAEGL S.A., GRIFFEY S.M., THEON A.P., MADEWELL B.R.;
RL          CANCER LETT. 107:125-130(1996).
RN          [2]
RP          SEQUENCE OF 76-280 FROM N.A.
RX          MEDLINE: 96293865.
RA          NASIR L., REID S.W.;

```

RL DNA SEQ. 6:185-187(1996).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
DR EMBL: S83123; G1836145; -.  
DR EMBL: U37120; G189675; -.  
DR PROSITE: PS00348; P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 1 274 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT NON\_TER 1 1  
FT CONFLICT 79 79 T -> A (IN REF. 2).  
FT CONFLICT 83 83 L -> M (IN REF. 2).  
FT CONFLICT 111 111 A -> V (IN REF. 2).  
FT CONFLICT 138 138 G -> A (IN REF. 2).  
FT NON\_TER 280 280  
SQ SEQUENCE 280 AA; 30985 MW; B494F872 CRC32;  
Query Match 89.2%; Score 66; DB 1; Length 280;  
Best Local Similarity 90.0%; Pred. No. 1.05e-01;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
DB 34 APAPAPSWPL 43  
QY 1 APAPAPSWPL 10  
RESULT 5  
ID P53-RABBIT STANDARD: PRT; 391 AA.  
AC Q95330;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS ORCOTOLAGUS CUNICULUS (RABBIT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; LAGOMORPHA.  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-NEW ZEALAND;  
RX MEDLINE: 9720869.  
RA LE GOS F., MAY P., RONCO P., CARON DE FROMENTEL C.;  
RL GENE 185:169-173(1997).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
DR EMBL: X90592; E194962; -.  
DR EMBL: X90592; E194962; -.  
DR PROSITE: PS00348; P53; 1.

KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 1 70 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 308 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 391 AA; 43435 MW; 30A36172 CRC32;  
Query Match 89.2%; Score 66; DB 1; Length 391;  
Best Local Similarity 90.0%; Pred. No. 1.05e-01;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
DB 81 APAPAPSWPL 90  
QY 1 APAPAPSWPL 10  
RESULT 6  
ID P53-BOVIN STANDARD: PRT; 386 AA.  
AC Q29628;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS BOS TAURUS (BOVINE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; ARTIODACTYLA.  
RN [1]  
RN SEQUENCE FROM N.A.  
RC TISSUE-LIVER;  
RX MEDLINE: 95352829.  
RA DEQUIEDT F., KERTMANN R., BURNY A., WILLEMS L.;  
RL DNA SEQ. 5:261-264(1995).  
RN [2]  
RN SEQUENCE OF 13-386 FROM N.A.  
RC STRAIN-HOLSTEIN; TISSUE-THYMUS;  
RX MEDLINE: 96401400.  
RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINGAWA M., AIDA Y.;  
RL VET. IMMUNOL. IMMUNOPATHOL. 52:53-63(1996).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
DR EMBL: X81704; G602333; -.  
DR EMBL: D49825; G1729419; -.  
DR PROSITE: PS00348; P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
FT CONFLICT 380 380 R -> T (IN REF. 2).  
SQ SEQUENCE 386 AA; 43255 MW; 0322BF3D CRC32;  
Query Match 86.5%; Score 64; DB 1; Length 386;  
Best Local Similarity 80.0%; Pred. No. 2.31e-01;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
DB 76 TPAPAPSWPL 85  
QY 1 APAPAPSWPL 10

RESULT 7  
ID P53 MOUSE STANDARD: PRT: 390 AA.  
AC P02340;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
EUTHERIA; RODENTIA.  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85027173.  
RA BIENZ B., ZAKUT-HOURI R., GIOVOL D., OREN M.,  
EMBO J. 3:2179-2183(1984).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 84068204.  
RA ZAKUT-HOURI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIOVOL D.;  
NATURE 306:594-597(1983).  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 84272240.  
RA JENKINS J.R., RUDGE K., REDMOND S., MADE-EVANS A.;  
NUCLEIC ACIDS RES. 12:5609-5626(1984).  
[4]  
RP SEQUENCE FROM N.A. (CLONES PCD53; P53-M11 AND P53-M8).  
RX MEDLINE: 87064640.  
RA ARAI N., NOMURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.,  
ROTTER V.;  
MOL. CELL. BIOL. 6:3232-3239(1986).  
[5]  
RP SEQUENCE OF 222-258 FROM N.A.  
RX MEDLINE: 92115342.  
RA BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BRENNER R.,  
BALMAIN A.;  
ONCOGENE 6:2363-2369(1991).  
[6]  
RP PHOSPHORYLATION SITES.  
RX MEDLINE: 86149247.  
RA SAMAD A., ANDERSON C.W., CARROLL R.B.;  
PROC. NATL. ACAD. SCI. U.S.A. 83:897-901(1986).  
[7]  
RP PHOSPHORYLATION SITES.  
RX MEDLINE: 91006019.  
RA MEER D.W., SIMON S., KIRKAWA U., ECKHART W.;  
EMBO J. 9:3253-3260(1990).  
[8]  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
DR EMBL: X00885; G871421; JOINED.  
DR EMBL: X00876; G871421; JOINED.  
DR EMBL: X00877; G871421; JOINED.  
DR EMBL: X00878; G871421; JOINED.  
DR EMBL: X00879; G871421; JOINED.  
DR EMBL: X00880; G871421; JOINED.  
DR EMBL: X00881; G871421; JOINED.  
DR EMBL: X00882; G871421; JOINED.  
DR EMBL: X00883; G871421; JOINED.  
DR EMBL: X00884; G871421; JOINED.

DR EMBL: X00885; G871421; JOINED.  
DR EMBL: K01700; G200205; -.  
DR EMBL: X01237; G53576; -.  
DR EMBL: X00741; G53571; -.  
DR EMBL: M13872; G200199; -.  
DR EMBL: M13873; G200201; -.  
DR EMBL: M13874; G200203; -.  
DR EMBL: S77930; G243255; -.  
DR PIR: A02684; DNMS53.  
DR PIR: A22739; A22739.  
DR PIR: S38822; S38822.  
DR HSSP: P04637; 1PES.  
DR TRANSFAC: T01806; -.  
DR MGD: MGI:98834; TRP53.  
DR PROSITE: PS00348; P53; 1.  
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS; DISEASE MUTATION.  
FT DOMAIN 1 75 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 76 150 HYDROPHOBIC.  
FT DOMAIN 276 390 HIGHLY BASIC AND MAY BE INVOLVED IN  
INTERACTION WITH DNA.  
FT DOMAIN 308 320 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 312 312 PHOSPHORYLATION (BY CK2).  
FT MOD\_RES 389 389 PHOSPHORYLATION.  
FT VARIANT 135 135 A -> V (CAN COOPERATE WITH AN ACTIVATED  
RAS TO TRANSFORM FIBROBLASTS).  
FT VARIANT 168 168 E -> G (IN CLONE P53-M11).  
FT CONFLICT 48 48 Q -> R (IN REF. 3).  
FT CONFLICT 79 81 PVA -> QW (IN REF. 3).  
SQ SEQUENCE 390 AA; 43458 MW; 8943DD93 CRC32;  
  
Query Match 85.1%; Score 63; DB 1; Length 390;  
Best Local Similarity 80.0%; Pred. No. 3.41e-01;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
DB 81 APAPAPMPL 90  
QY 1 APAPAPMPL 10  
  
RESULT 8  
ID P53 SPEBE STANDARD: PRT: 314 AA.  
AC O64662;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53.  
OS SPERMOPHILUS BEECHERI (BEECHER GROUND SQUIRREL).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
EUTHERIA; RODENTIA.  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 95007566.  
RA RIVKIN M.B., COLLEN J.M., ROBINSON W.S., MARION P.L.;  
CANCER RES. 54:5430-5437(1994).  
[2]  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
DR EMBL: U43902; G1165312; -.

DR PROSITE: PS00348; P53; 1.  
 KM ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT NON-TER 1 1 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 289 301  
 FT NON-TER 314 314  
 SQ SEQUENCE 314 AA; 34618 MW; D07F433B CRC32;

Query Match 81.1%; Score 60; DB 1; Length 314;  
 Best Local Similarity 80.0%; Pred. No. 1.08e+00;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 62 APAPASWPL 71  
 1111111111  
 1 APAPASWPL 10

RESULT 9  
 ID P53\_FELCA STANDARD; PRT; 386 AA.  
 AC P41685;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS FELIS SILVESTRIUS CATUS (CAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; CARNIVORA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LYMPH NODE;  
 RX MEDLINE; 94333960.  
 RA OKUDA M., UMEIDA A., SAKAI T., OHASHI T., MOMOI Y., YOUN H.Y.,  
 RA WATABE T., GOITSUKA R., TSUJIMOTO H., HASEGAWA A.;  
 RL INT. J. CANCER 58:602-607(1994).  
 RN [2]  
 RP SEQUENCE OF 34-354 FROM N.A.  
 RX MEDLINE; 94114699.  
 RA OKUDA M., UMEIDA A., MATSUMOTO Y., MOMOI Y., WATABE T., GOITSUKA R.,  
 RA O'BRIEN S.J., TSUJIMOTO H., HASEGAWA A.;  
 RL J. VET. MED. SCI. 55:801-805(1993).

CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.

DR EMBL; D26608; G538225; -;  
 DR EMBL; D16460; G575528; -;  
 DR PROSITE; PS00348; P53; 1.  
 KM ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD-RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 285 285 K -> R (IN REF. 2).  
 SQ SEQUENCE 386 AA; 42692 MW; D6C7132A CRC32;

Query Match 81.1%; Score 60; DB 1; Length 386;  
 Best Local Similarity 80.0%; Pred. No. 1.08e+00;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 76 TPAPASWPL 85  
 1111111111

OY 1 APAPASWPL 10

RESULT 10  
 ID GND5\_RAT STANDARD; PRT; 895 AA.  
 AC G03386;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
 DE GUANINE NUCLEOTIDE DISSOCIATION STIMULATOR RALGDSB.  
 DE RATTUS NORVEGICUS (RAT).  
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-FIBROBLAST;  
 RX MEDLINE; 93154339.  
 RA ALBRICHT C.F., GIDDINGS B.W., LIU J., VINO M., WEINBERG R.A.;  
 RL EMBL J. 12:339-347(1993).  
 CC -1- FUNCTION: STIMULATES THE DISSOCIATION OF GDP FROM THE RAS-RELATED  
 CC RALA AND RALB GTPASES WHICH ALLOWS GTP BINDING AND ACTIVATION OF  
 CC THE GTPASES.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED.  
 CC -1- SIMILARITY: TO OTHER GUANINE-NUCLEOTIDE RELEASING FACTORS OF THE  
 CC CDC25 FAMILY.  
 DR EMBL; I07925; G204438; -;  
 DR PROSITE; PS00720; GDS.CDC25; 1.  
 DR PHOSPHORYLATION; GUANINE-NUCLEOTIDE RELEASING FACTOR.  
 SQ SEQUENCE 895 AA; 98869 MW; B8F60F3C CRC32;

Query Match 78.4%; Score 58; DB 1; Length 895;  
 Best Local Similarity 80.0%; Pred. No. 2.27e+00;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 340 APAPASWPL 349  
 1111111111  
 1 APAPASWPL 10

RESULT 11  
 ID P53\_RAT STANDARD; PRT; 391 AA.  
 AC P10361; G09168;  
 DT 01-MAR-1989 (REL. 10, CREATED)  
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR P53.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE; 89083585.  
 RA SOUSSEI T.;  
 RL NUCLEIC ACIDS RES. 16:11384-11384(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE; 93181268.  
 RA HULLA J.E., SCHNEIDER R.P.;  
 RL NUCLEIC ACIDS RES. 21:713-717(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RA MATHOPALA S.P.;  
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF



CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: X13058; G56829; -  
 DR EMBL: L07910; G205952; -  
 DR EMBL: L07904; G205952; JOINED.  
 DR EMBL: L07905; G205952; JOINED.  
 DR EMBL: L07906; G205952; JOINED.  
 DR EMBL: L07907; G205952; JOINED.  
 DR EMBL: L07908; G205952; JOINED.  
 DR EMBL: L07909; G205952; JOINED.  
 DR EMBL: U90328; G1938365; -  
 DR PIR: S02192; S02192.  
 DR HSSP: P04637; 1PES.  
 DR PROSITE: PS00348; P53; 1.  
 DR ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 76 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 77 151 HIGHLY BASIC AND MAY BE INVOLVED IN  
 FT DOMAIN 277 391 INTERACTION WITH DNA.  
 FT DOMAIN 309 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).  
 FT VARIANT 103 103 G -> S.  
 FT VARIANT 256 256 E -> G.  
 FT CONFLICT 174 174 C -> W (IN REF. 2).  
 SQ SEQUENCE 391 AA; 43451 MM; E0114C18 CRC32;.

Query Match 75.7%; Score 56; DB 1; Length 391;  
 Best Local Similarity 70.0%; Pred. No. 4.75e+00;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 82 APASATPMP 91  
 QY 1 APADAPSWPL 10

RESULT 12  
 ID P53\_CRIGR STANDARD: PRT: 393 AA.  
 AC 009185; G64397; P97258; P97788;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR P53.  
 OS CRICETUS GRISEUS (CHINESE HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RA CHAUNG W., MI L.J., BOORSTEIN R.J.;  
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE-LIVER;  
 RC MEDLINE: 97183659.  
 RX LEE H., LARNER J.M., HAMLIN J.L.;  
 RL GENE 184:177-183(1997).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: Y08900; E303876; -  
 DR EMBL: Y08901; E303863; -  
 DR EMBL: U50395; G1842230; -  
 DR PROSITE: PS00348; P53; 1.  
 DR ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 74 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 75 150 HIGHLY BASIC AND MAY BE INVOLVED IN  
 FT DOMAIN 316 390 INTERACTION WITH DNA.  
 FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
 FT VARIANT 133 133 L -> Q (IN CELL LINE V79-4).  
 FT VARIANT 135 135 C -> W (IN CELL LINE V79-4).  
 FT CONFLICT 103 103 Y -> F (IN REF. 2).  
 SQ SEQUENCE 393 AA; 43378 MM; 402EB149 CRC32;.

Query Match 75.7%; Score 56; DB 1; Length 393;  
 Best Local Similarity 70.0%; Pred. No. 4.75e+00;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 84 ASAPATPMP 93  
 QY 1 APADAPSWPL 10

RESULT 13  
 ID P53\_MESAU STANDARD: PRT: 396 AA.  
 AC 000366; P97276;  
 DT 01-DEC-1992 (REL. 24, CREATED)  
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS MESOCRICETUS AURATUS (GOLDEN HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RA STRAIN-SYRIAN; TISSUE-KIDNEY;  
 RC MEDLINE: 92210007.  
 RA LEGROS Y., MCINTYRE P., SOUSSI T.;  
 RL GENE 112:247-250(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP HOU E.W., WISEMAN R.;  
 RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: M75144; G191415; -  
 DR EMBL: U07182; G473579; -  
 DR PIR: JH0633; JH0633.  
 DR HSSP: P04637; 1PES.  
 DR PROSITE: PS00348; P53; 1.  
 DR ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.

[illegible]

FT DOMAIN 2209

FT DOMAIN 2209

Sun Sep 13 10:55:30 1998

|    |          |          |            |  |
|----|----------|----------|------------|--|
| FT | NP-BIND  | 2215     | 2223       | ATP (BY SIMILARITY).                     |
| FT | BINDING  | 2242     | 2242       | ATP (BY SIMILARITY).                     |
| FT | MUTAGEN  | 2242     | 2242       | K->M: INACTIVATES THE PROTEIN.           |
| FT | MOD RES  | 2380     | 2380       | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). |
| FT | CARBOHYD | 30       | 30         | POTENTIAL.                               |
| FT | CARBOHYD | 129      | 129        | POTENTIAL.                               |
| FT | CARBOHYD | 481      | 481        | POTENTIAL.                               |
| FT | CARBOHYD | 505      | 505        | POTENTIAL.                               |
| FT | CARBOHYD | 617      | 617        | POTENTIAL.                               |
| FT | CARBOHYD | 647      | 647        | POTENTIAL.                               |
| FT | CARBOHYD | 966      | 966        | POTENTIAL.                               |
| FT | CARBOHYD | 1228     | 1228       | POTENTIAL.                               |
| FT | CARBOHYD | 1313     | 1313       | POTENTIAL.                               |
| FT | CARBOHYD | 1353     | 1353       | POTENTIAL.                               |
| FT | CARBOHYD | 1550     | 1550       | POTENTIAL.                               |
| FT | CARBOHYD | 1557     | 1557       | POTENTIAL.                               |
| FT | CARBOHYD | 1639     | 1639       | POTENTIAL.                               |
| FT | CARBOHYD | 1725     | 1725       | POTENTIAL.                               |
| FT | CARBOHYD | 1756     | 1756       | POTENTIAL.                               |
| FT | CARBOHYD | 1804     | 1804       | POTENTIAL.                               |
| FT | CARBOHYD | 1889     | 1889       | POTENTIAL.                               |
| FT | CARBOHYD | 1947     | 1947       | POTENTIAL.                               |
| FT | CARBOHYD | 2073     | 2073       | POTENTIAL.                               |
| FT | VARIANT  | 392      | 392        | M->V.                                    |
| FT | VARIANT  | 1668     | 1668       | A->V.                                    |
| FT | VARIANT  | 1703     | 1703       | N->H.                                    |
| FT | VARIANT  | 1730     | 1730       | R->K.                                    |
| FT | VARIANT  | 1731     | 1731       | G->E.                                    |
| FT | VARIANT  | 1741     | 1741       | V->M.                                    |
| FT | VARIANT  | 2271     | 2271       | R->C.                                    |
| FT | CONFLICT | 1823     | 1823       | E->Q (IN REF. 2).                        |
| SO | SEQUENCE | 2554 AA; | 287107 MW; | 1143D891 CRC32;                          |

Query Match 75.7%; Score 56; DB 1; Length 2554;  
Best Local Similarity 60.0%; Pred. No. 4,75e+00;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 793 AASPSASWPL 802

QY 1 APAPAPSWPL 10

Search completed: Fri Sep 11 13:29:41 1998  
Job time : 5 secs.

**THIS PAGE BLANK (USPTO)**

(TM)

Distribution rights by Oxford Molecular Ltd

```
Run on:      Fri Sep 11 13:29:59 1998; MasPar time 4.19 Seconds
```

not generated.

Description: (1-10) FROM US08432843.pdf  
 Perfect Score: 74

Scoring table: PAM 150

```
post-processing: Minimum Match 0%
Listing first 45 summaries
```

1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mmc 6:sp\_organelle 7:sp\_phage 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp Vertebrate  
13:sp\_unclassified

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

| Result | Query | DB    | ID     | Description | Pred. No |
|--------|-------|-------|--------|-------------|----------|
| No.    | Score | Match | Length |             |          |

| No. | Score | Match Length | DB   | ID | Description | Pred. No.              |          |
|-----|-------|--------------|------|----|-------------|------------------------|----------|
| 1   | 74    | 100.0        | 245  | 2  | 015085      | P53 TRANSFORMATION SUP | 8.23e-03 |
| 2   | 74    | 100.0        | 393  | 2  | 015809      | CELLULAR TUMOR ANTIGEN | 8.23e-03 |
| 3   | 74    | 100.0        | 393  | 2  | 016811      | CELLULAR TUMOR ANTIGEN | 8.23e-03 |
| 4   | 74    | 100.0        | 393  | 2  | 016807      | CELLULAR TUMOR ANTIGEN | 8.23e-03 |
| 5   | 74    | 100.0        | 393  | 2  | 016808      | CELLULAR TUMOR ANTIGEN | 8.23e-03 |
| 6   | 74    | 100.0        | 393  | 2  | 015087      | CELLULAR TUMOR ANTIGEN | 8.23e-03 |
| 7   | 74    | 100.0        | 393  | 2  | 015080      | P53 TRANSFORMATION SUP | 8.23e-03 |
| 8   | 74    | 100.0        | 393  | 2  | 016810      | CELLULAR TUMOR ANTIGEN | 8.23e-03 |
| 9   | 74    | 100.0        | 393  | 2  | 016848      | CELLULAR TUMOR ANTIGEN | 8.23e-03 |
| 10  | 74    | 100.0        | 393  | 2  | 016535      | P53 TRANSFORMATION SUP | 8.23e-03 |
| 11  | 74    | 100.0        | 393  | 2  | 015086      | P53 TRANSFORMATION SUP | 8.23e-03 |
| 12  | 71    | 95.9         | 285  | 4  | 095326      | CELLULAR TUMOR ANTIGEN | 2.78e-02 |
| 13  | 64    | 86.5         | 391  | 13 | 036006      | CELLULAR TUMOR ANTIGEN | 4.38e-01 |
| 14  | 63    | 85.1         | 378  | 11 | P83002      | P53 (FRAGMENT)         | 6.42e-01 |
| 15  | 56    | 75.1         | 837  | 9  | 026752      | ATP-DEPENDENT RNA HELI | 8.68e+00 |
| 16  | 56    | 75.7         | 2559 | 3  | 024512      | SEVENLESS MRNA.        | 1.24e+01 |
| 17  | 55    | 74.3         | 814  | 2  | 013493      | MDG15.                 | 1.24e+01 |
| 18  | 55    | 74.3         | 814  | 2  | 013444      | METARGIDIN PRECURSOR.  | 1.24e+01 |
| 19  | 55    | 74.0         | 1937 | 9  | 030482      | PKS MODULE 4.          | 1.77e+01 |
| 20  | 54    | 73.0         | 224  | 2  | 095722      | TISSUE INHIBITOR OF ME | 1.77e+01 |

|    |    |      |      |    |        |                         |          |
|----|----|------|------|----|--------|-------------------------|----------|
| 21 | 53 | 71.6 | 168  | 11 | Q79748 | NEF PROTEIN.            | 2.52e+01 |
| 22 | 53 | 71.6 | 206  | 11 | Q79691 | NEF PROTEIN.            | 2.52e+01 |
| 23 | 53 | 71.6 | 206  | 11 | Q79680 | NEF PROTEIN.            | 2.52e+01 |
| 24 | 53 | 71.6 | 206  | 11 | Q72426 | NEF PROTEIN.            | 2.52e+01 |
| 25 | 53 | 71.6 | 206  | 11 | Q79753 | NEF PROTEIN.            | 2.52e+01 |
| 26 | 53 | 71.6 | 206  | 11 | Q79755 | NEF PROTEIN.            | 2.52e+01 |
| 27 | 53 | 71.6 | 206  | 11 | Q79694 | NEF PROTEIN.            | 2.52e+01 |
| 28 | 53 | 71.6 | 206  | 11 | Q79693 | NEF PROTEIN.            | 2.52e+01 |
| 29 | 53 | 71.6 | 206  | 11 | Q72427 | NEF PROTEIN.            | 2.52e+01 |
| 30 | 53 | 71.6 | 206  | 11 | Q79749 | NEF PROTEIN.            | 2.52e+01 |
| 31 | 53 | 71.6 | 206  | 11 | Q79747 | NEF PROTEIN.            | 2.52e+01 |
| 32 | 53 | 71.6 | 206  | 11 | Q79746 | NEF PROTEIN.            | 2.52e+01 |
| 33 | 53 | 71.6 | 206  | 11 | Q79721 | NEF PROTEIN.            | 2.52e+01 |
| 34 | 53 | 71.6 | 206  | 11 | Q79740 | NEF PROTEIN.            | 2.52e+01 |
| 35 | 53 | 71.6 | 206  | 11 | Q79741 | NEF PROTEIN.            | 2.52e+01 |
| 36 | 53 | 71.6 | 206  | 11 | Q79719 | NEF PROTEIN.            | 2.52e+01 |
| 37 | 53 | 71.6 | 206  | 11 | Q79716 | NEF PROTEIN.            | 2.52e+01 |
| 38 | 53 | 71.6 | 206  | 11 | Q79744 | NEF PROTEIN.            | 2.52e+01 |
| 39 | 53 | 71.6 | 206  | 11 | Q79745 | NEF PROTEIN.            | 2.52e+01 |
| 40 | 53 | 71.6 | 206  | 11 | Q79710 | NEF PROTEIN.            | 2.52e+01 |
| 41 | 53 | 71.6 | 207  | 11 | Q79725 | NEF PROTEIN (FRAGMENT)  | 2.52e+01 |
| 42 | 53 | 71.6 | 495  | 9  | Q48384 | PLASMD R751 GENES FRO   | 2.52e+01 |
| 43 | 53 | 71.6 | 670  | 2  | Q14560 | DISHELLED 1.            | 2.52e+01 |
| 44 | 53 | 71.6 | 903  | 2  | Q14560 | PSY965_3                | 2.52e+01 |
| 45 | 53 | 71.6 | 4340 | 9  | Q30764 | POLYPEPTIDE SYNTHASE MO | 2.52e+01 |

|              |        |
|--------------|--------|
| RESULT       | 1      |
| ID           | Q15085 |
| PRELIMINARY; | PRT;   |
|              | 245 AA |

```

DN      01-NOV-1996 (TREMBLERL. 01, CREATED)
DT      01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)
DT      01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)
DT      01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)
DE      P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN      P53.
OC      HOMO SAPIENS (HUMAN).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.
OC      EUHERIA; PRIMATES.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 92007731.
RA      FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.:
RL      EMBO J. 10:2879-2887(1991).
DR      EMBL; X60010: G506433: -.
FT      NON_TER      245      245
SQ      SEQUENCE      245 AA; 27066 MW; 55B80C07 CRC32;

Query Match      100.0%; Score 74; DB 2; Length 245;
Best Local Similarity 100.0%; Pred. No. 8.23e-03;
Matches      10; Conservative      0; Mismatches      0; Indels      0;

DB      84 APAPASWPL 93
      ||||||||
QY      1 APAPASWPL 10

RESULT      2
ID      Q16809      PRELIMINARY; PRT; 393 AA.
AC      Q16809.
DT      01-NOV-1996 (TREMBLERL. 01, CREATED)
DT      01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)
DT      01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)
DE      CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN      P53.
OS      HOMO SAPIENS (HUMAN).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.
OC      EUHERIA; PRIMATES.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 92007731.
RA      FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.:

```

```

RL  EMO J. 10:2879-2887(1991).
CC  -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC  PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC  CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC  REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC  FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC  CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC  -1- SUBCELLULAR LOCATION: NUCLEAR.
DR  EMBL: X60019: G506451: -
DR  PROSITE: PS00348: P53; 1
KW  ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW  NUCLEAR PROTEIN; PHOSPHORYLATION.
FT  VARIANT: 213 213 Q -> R.
FT  NON_TER: 393 393
SQ  SEQUENCE 393 AA: 43684 MW: CB70BD7F CRC32:

Query Match 100.0%; Score 74; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 8.23e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPASWPL 93
OY 1 APAPASWPL 10

RESULT 3
ID 016811: PRELIMINARY; PRT; 393 AA.
AC 016811:
DT 01-NOV-1996 (TREMBLREL, 01, CREATED)
DT 01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL, 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 85126934.
RA MATIASHEWSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,
RA BENCHIMOL S.;
RA EMO J. 3:3237-3262(1984).
NN (12)
RP SEQUENCE FROM N.A.
RX MEDLINE: 87064416.
RA LAMB P., CRAWFORD L.;
RL MOL. CELL. BIOL. 6:1379-1385(1986).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR  EMBL: M13121: G386994: JOINED.
DR  EMBL: M13112: G386994: JOINED.
DR  EMBL: M13113: G386994: JOINED.
DR  EMBL: M13114: G386994: JOINED.
DR  EMBL: M13115: G386994: JOINED.
DR  EMBL: M13116: G386994: JOINED.
DR  EMBL: M13117: G386994: JOINED.
DR  EMBL: M13118: G386994: JOINED.
DR  EMBL: M13119: G386994: JOINED.
DR  EMBL: M13120: G386994: JOINED.
DR  EMBL: M13120: G386994: JOINED.
DR  PROSITE: PS00348: P53; 1.
KW  REPEAT; TUMOR ANTIGEN; ANTI-ONCOGENE; DNA-BINDING;
KW TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN;
KW PHOSPHORYLATION.
FT  NON_TER: 393 393
SQ  SEQUENCE 393 AA: 43698 MW: 3EA71431 CRC32:

Query Match 100.0%; Score 74; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 8.23e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 84 APAPASWPL 93
OY 1 APAPASWPL 10

RESULT 4
ID 016807: PRELIMINARY; PRT; 393 AA.
AC 016807:
DT 01-NOV-1996 (TREMBLREL, 01, CREATED)
DT 01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL, 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RA EMO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR  EMBL: X60018: G506449: -
DR  PROSITE: PS00348: P53; 1.
KW  ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW NUCLEAR PROTEIN; PHOSPHORYLATION.
SQ  SEQUENCE 393 AA: 43731 MW: 279BC9CB CRC32:

Query Match 100.0%; Score 74; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 8.23e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPASWPL 93
OY 1 APAPASWPL 10

RESULT 5
ID 016808: PRELIMINARY; PRT; 393 AA.
AC 016808:
DT 01-NOV-1996 (TREMBLREL, 01, CREATED)
DT 01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL, 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RA EMO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR  EMBL: X60018: G506449: -
DR  PROSITE: PS00348: P53; 1.
KW  ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW NUCLEAR PROTEIN; PHOSPHORYLATION.
SQ  SEQUENCE 393 AA: 43731 MW: 279BC9CB CRC32:

```

FT VARIANT 163 163 H -> Y.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43627 MW: AFDBA9E3 CRC32:

Query Match 100.0%; Score 74; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8,23e-03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 84 AAPAPSWPL 93  
 1 AAPAPSWPL 10

QY 1 AAPAPSWPL 10

RESULT 6 PRELIMINARY; PRT; 393 AA.

ID 015087;  
 AC 015087;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.

RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL: X60014; G506441; -;  
 FT VARIANT 237 237 I -> M.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43694 MW: 9BB81992 CRC32:

Query Match 100.0%; Score 74; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8,23e-03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 84 AAPAPSWPL 93  
 1 AAPAPSWPL 10

QY 1 AAPAPSWPL 10

RESULT 7 PRELIMINARY; PRT; 393 AA.

ID 015088;  
 AC 015088;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.

RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL: X60016; G506445; -;  
 FT VARIANT 238 238 Y -> C.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43713 MW: A01E1523 CRC32:

Query Match 100.0%; Score 74; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8,23e-03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 84 AAPAPSWPL 93  
 1 AAPAPSWPL 10

QY 1 AAPAPSWPL 10

RESULT 8 PRELIMINARY; PRT; 393 AA.

ID 016810;  
 AC 016810;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).

OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.

RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: X60020; G506453; -;  
 DR PROSITE: PS00348; P53: 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 254 254 D -> N.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43714 MW: 5F914579 CRC32:

Query Match 100.0%; Score 74; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8,23e-03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 84 AAPAPSWPL 93  
 1 AAPAPSWPL 10

QY 1 AAPAPSWPL 10

RESULT 9 PRELIMINARY; PRT; 393 AA.

ID 016848;  
 AC 016848;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.

OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.

RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER W., WOLF D., ARAI N.,  
 RA ROTTER V.;  
 RL MOL. CELL. BIOL. 6:4650-4656(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: M14694; G339814; -;  
 DR PROSITE: PS00348; P53: 1.  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ANTI-ONCOGENE; DNA-BINDING;  
 KW TRANSCRIPTION REGULATION; ACTIVATOR.

Query Match 100.0%; Score 74; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8,23e-03;

DB 84 AAPAPSWPL 93  
 1 AAPAPSWPL 10

QY 1 AAPAPSWPL 10

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPASWPL 93

OY 1 APAPASWPL 10

RESULT 10 PRELIMINARY: PRT: 393 AA.

ID 016535 AC 016535.

DT 01-NOV-1996 (TREMUREL. 01, LAST SEQUENCE UPDATE)

DE 01-NOV-1996 (TREMUREL. 01, LAST SEQUENCE UPDATE)

DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

GN P53.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 92007731.

RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;

RL EMO J. 10:2879-2887(1991).

DR EMBL: X60017; G506443; -

FT VARIANT 248 248 Q -> R.

FT NON TER 393 393

SO SEQUENCE 393 AA; 43684 MW; 239818A9 CRC32;

Query Match  
Best Local Similarity 100.0%; Score 74; DB 2; Length 393;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPASWPL 93

OY 1 APAPASWPL 10

RESULT 11 PRELIMINARY: PRT: 393 AA.

ID 015086 AC 015086.

DT 01-NOV-1996 (TREMUREL. 01, LAST SEQUENCE UPDATE)

DE 01-NOV-1996 (TREMUREL. 01, LAST SEQUENCE UPDATE)

DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

GN P53.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 92007731.

RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;

RL EMO J. 10:2879-2887(1991).

DR EMBL: X60013; G506443; -

FT VARIANT 246 246 T -> M.

FT NON TER 393 393

SO SEQUENCE 393 AA; 43682 MW; 943B62A3 CRC32;

Query Match  
Best Local Similarity 100.0%; Score 74; DB 2; Length 393;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPASWPL 93

OY 1 APAPASWPL 10

RESULT 12 PRELIMINARY: PRT: 285 AA.

ID 095326 AC 095326.

DT 01-FEB-1997 (TREMUREL. 02, CREATED)

DT 01-FEB-1997 (TREMUREL. 02, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMUREL. 05, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).

GN P53.

OS CANIS FAMILIARIS (DOG).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; CARNIVORA.

GN [1]

RP SEQUENCE FROM N.A.

RA YANG B.J., SHI X.B., LAU D.H.M.;

RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT

CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL

CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY

CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED

CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF

CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL: U62133; G1619833; -

DR PROSITE: PS00348; P53; 1.

KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;

KW NUCLEAR PROTEIN; PHOSPHORYLATION.

FT NON TER 1 1

FT NON TER 285 285

SO SEQUENCE 285 AA; 31616 MW; 15E1EC47 CRC32;

Query Match  
Best Local Similarity 95.9%; Score 71; DB 4; Length 285;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 39 APAPASWPL 48

OY 1 APAPASWPL 10

RESULT 13 PRELIMINARY: PRT: 391 AA.

ID 036006 AC 036006.

DT 01-JAN-1998 (TREMUREL. 05, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMUREL. 05, LAST SEQUENCE UPDATE)

DE 01-JAN-1998 (TREMUREL. 05, LAST ANNOTATION UPDATE)

DE CELLULAR TUMOR ANTIGEN P53.

GN P53.

OS MAMMATA MONAX.

OC PLASMID P7BLUE (R).

OC UNCLASSIFIED.

GN [1]

RP SEQUENCE FROM N.A.

RA FEITELSON M.A., RANGANTHAN P.N., CLAYTON M.M., ZHANG S.M.;

RL ONCOGENE 15:327-336(1997).

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT

CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL

CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY

CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED

CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF

CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL: AJ001022; E351287; -

DR PROSITE: PS00346; P53; 1.

KW PLASMID; ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION;

KW ACTIVATOR; NUCLEAR PROTEIN; PHOSPHORYLATION.

SO SEQUENCE 391 AA; 43468 MW; 95FAB8F2 CRC32;

Query Match  
Best Local Similarity 86.5%; Score 64; DB 13; Length 391;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 82 ASPAPSWPL 91

OY 1 APAPASWPL 10

RESULT 14 PRELIMINARY: PRT: 378 AA.

ID P89002 AC P89002.



AC P89002;  
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)  
 DE P53 (FRAGMENT)  
 OS MASTOMYS NATALENSIS PAPILLOMAVIRUS (MNPV).  
 OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPAPOVIRIDAE; PAPILLOMAVIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA LUQUE E.A., TANG L.H., MODLIN I.M.;  
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U48616; FGI813451;  
 FT NON\_TER  
 SQ SEQUENCE 378 AA; 42062 MW; B4436760 CRC32;

Query Match 85.18; Score 63; DB 11; Length 378;  
 Best Local Similarity 80.08; Pred. No. 6.42e-01;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 69 APAPAPMPL 78  
 |||||  
 QY 1 APAPAPMPL 10

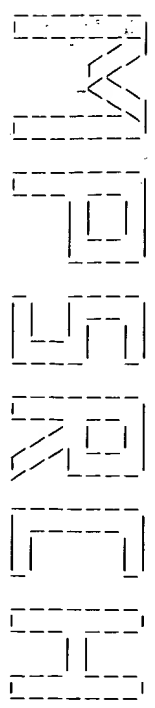
RESULT 15  
 ID 026752 PRELIMINARY; PRT: 837 AA.  
 AC 026752;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE ATP-DEPENDENT RNA HELICASE RELATED PROTEIN.  
 GN MTH656.  
 OS METHANOBACTERIUM THERMAUTOTROPHICUM.  
 OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOBACTERIALES;  
 OC METHANOBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DELTA H;  
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DOBOIS J.,  
 RA ALBREDE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,  
 RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTIER B., QIU D.,  
 RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,  
 RA JIMANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,  
 RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,  
 RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;  
 RL J. BACTERIOL. 179:7135-7155(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DELTA H;  
 RA SMITH D.R.;  
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AE000845; G2621738;  
 KW HELICASE.  
 SQ SEQUENCE 837 AA; 93979 MW; 27C34F57 CRC32;

Query Match 75.78; Score 56; DB 9; Length 837;  
 Best Local Similarity 75.08; Pred. No. 8.68e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 613 SPDPSPMPL 620  
 :|||||  
 QY 3 APAPSPMPL 10

Search completed: Fri Sep 11 13:30:45 1998  
 Job time : 46 secs.

**THIS PAGE BLANK (USPTO)**



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSearch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:24:53 1998; Maspar time 2.75 Seconds  
58.833 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-13  
Description: (1-10) from US08452843. pep  
Perfect Score: 67  
Sequence: 1 LPENNVLSPL 10

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 16.051; Variance 52.131; scale 0.308

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID  | Description            | Pred. No. |
|------------|-------|-------------|--------|--------|------------------------|-----------|
| 1          | 67    | 100.0       | 11 20  | R97509 | Cytotoxic T lymphocyte | 1.22e+00  |
| 2          | 67    | 100.0       | 15 10  | R54911 | Immunodominant epitope | 1.22e+00  |
| 3          | 67    | 100.0       | 23 10  | R34907 | Immunodominant epitope | 1.22e+00  |
| 4          | 67    | 100.0       | 36 25  | W13604 | p53 protein amino acid | 1.22e+00  |
| 5          | 67    | 100.0       | 41 25  | W13603 | p53 protein amino acid | 1.22e+00  |
| 6          | 67    | 100.0       | 52 25  | W13602 | p53 protein amino acid | 1.22e+00  |
| 7          | 67    | 100.0       | 64 28  | W42878 | N-terminal region of   | 1.22e+00  |
| 8          | 67    | 100.0       | 64 28  | W42873 | Human p53 fragment 13  | 1.22e+00  |
| 9          | 67    | 100.0       | 64 28  | W42870 | N-terminal region of   | 1.22e+00  |
| 10         | 67    | 100.0       | 64 19  | W07886 | Human p53, involved in | 1.22e+00  |
| 11         | 67    | 100.0       | 71 28  | W47079 | Human p53 fragment (r  | 1.22e+00  |
| 12         | 67    | 100.0       | 241 10 | R51872 | Human p53 amino acids  | 1.22e+00  |
| 13         | 67    | 100.0       | 355 22 | W13950 | Del356-393 modified h  | 1.22e+00  |
| 14         | 67    | 100.0       | 361 21 | W13958 | Chimeric p53 protein.  | 1.22e+00  |
| 15         | 67    | 100.0       | 361 21 | W13861 | Chimeric p53 protein.  | 1.22e+00  |
| 16         | 67    | 100.0       | 363 22 | W13954 | Modified p53 variant   | 1.22e+00  |
| 17         | 67    | 100.0       | 363 21 | W13971 | Modified p53 variant   | 1.22e+00  |
| 18         | 67    | 100.0       | 363 21 | W13974 | Modified p53 variant   | 1.22e+00  |

|    |    |       |        |        |                       |          |
|----|----|-------|--------|--------|-----------------------|----------|
| 19 | 67 | 100.0 | 363 21 | W13973 | Modified p53 variant  | 1.22e+00 |
| 20 | 67 | 100.0 | 363 21 | W13972 | Modified p53 variant  | 1.22e+00 |
| 21 | 67 | 100.0 | 368 21 | W13956 | Chimeric p53 protein. | 1.22e+00 |
| 22 | 67 | 100.0 | 370 21 | W13957 | Chimeric p53 protein. | 1.22e+00 |
| 23 | 67 | 100.0 | 393 24 | W25155 | Human p53 variant fou | 1.22e+00 |
| 24 | 67 | 100.0 | 393 22 | W13949 | T284R modified human  | 1.22e+00 |
| 25 | 67 | 100.0 | 393 22 | W13979 | Human tumour-derived  | 1.22e+00 |
| 26 | 67 | 100.0 | 393 22 | W13953 | T284K modified human  | 1.22e+00 |
| 27 | 67 | 100.0 | 393 22 | W13948 | Human wild-type p53 t | 1.22e+00 |
| 28 | 67 | 100.0 | 393 21 | W05345 | Human p53 mutant N239 | 1.22e+00 |
| 29 | 67 | 100.0 | 393 22 | W13951 | Human tumour-derived  | 1.22e+00 |
| 30 | 67 | 100.0 | 393 21 | W05344 | Human p53.            | 1.22e+00 |
| 31 | 67 | 100.0 | 393 22 | W13952 | Human tumour-derived  | 1.22e+00 |
| 32 | 67 | 100.0 | 393 22 | W13978 | Human tumour-derived  | 1.22e+00 |
| 33 | 67 | 100.0 | 393 22 | W13981 | Human tumour-derived  | 1.22e+00 |
| 34 | 67 | 100.0 | 393 22 | W02617 | Human p53 tumour supp | 1.22e+00 |
| 35 | 67 | 100.0 | 393 22 | W13980 | Human tumour-derived  | 1.22e+00 |
| 36 | 67 | 100.0 | 393 16 | R26758 | p53 protein.          | 1.22e+00 |
| 37 | 67 | 100.0 | 393 16 | R94623 | Human p53 mutant R248 | 1.22e+00 |
| 38 | 67 | 100.0 | 393 21 | W05347 | Wild type p53 protein | 1.22e+00 |
| 39 | 67 | 100.0 | 393 18 | R91933 | Modified p53 variant  | 1.22e+00 |
| 40 | 67 | 100.0 | 393 21 | W13969 | Chimeric p53 protein. | 1.22e+00 |
| 41 | 67 | 100.0 | 402 21 | W13965 | Chimeric p53 protein. | 1.22e+00 |
| 42 | 67 | 100.0 | 406 21 | W13966 | Chimeric p53 protein. | 1.22e+00 |
| 43 | 67 | 100.0 | 411 21 | W13967 | Chimeric p53 protein. | 1.22e+00 |
| 44 | 67 | 100.0 | 438 14 | R74272 | Tumour suppressor pro | 1.22e+00 |
| 45 | 67 | 100.0 | 533 23 | W19763 | p53-GW-CSF immunostim | 1.22e+00 |

ALIGNMENTS

RESULT 1  
ID R97509 standard; peptide: 11 AA.  
AC R97509:  
DE 11-EB-1997 (first entry)  
DT Cytotoxic T lymphocyte-activating peptide, corresp. to p53 aa 25-35.  
KW p53; Her-2; Neu; aa; amino acid; CTL; cytotoxic T lymphocyte; target;  
KW Homologous cell; antigenic; vaccine; immunisation; activation.  
OS Homo sapiens.  
PN W09618409-A1.  
PD 20-JUN-1996.  
PF 14-DEC-1995; U16415.  
PR 14-DEC-1994; US-355558.  
PA (Scri) SCRIPPS RES INSTR.  
PI Sherman LA.  
DR WPI; 96-300385/30.  
PT In vivo activation of tumour-specific cytotoxic T lymphocytes - by  
PT contacting with polypeptide(s) derived from human p53 or Her-2/Neu  
PS DCIaim 40; Page 73; 158pp; English.  
CC R97509 is a peptide capable of activating cytotoxic T lymphocytes  
CC (CTLs) which specifically target malignant cells. The peptide  
CC corresponds to amino acids 25-35 of human p53 protein. CTL-  
CC activating peptides can be used in a vaccine for protecting against  
CC tumour cell formation. CTLs activated by the peptides will lyse  
CC tumour cells displaying specific peptides. Antibodies against CTL-  
CC activating peptides are useful for the identification of other  
CC similar compounds which may be useful for treating cancer or virally-  
CC infected cells, or for diagnosis. The peptide and vaccines produced  
CC provide immunity to a high percentage of different ethnic groups,  
CC i.e. those with different HLA alleles.  
SQ Sequence 11 AA:  
Query Match 100.0%; Score 67; DB 20; Length 11;  
Best Local Similarity 100.0%; Pred. No. 1.22e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 2 lpennvlspl 11  
Oy 1 LPENNVLSPL 10  
RESULT 2

ID R54911 standard; peptide; 15 AA.  
 AC R54911;  
 DT 29-NOV-1994 (first entry)  
 DE Immunodominant epitope from p53 N-terminal.  
 KW cancer; pre-cancerous state; detection; diagnosis; human p53 gene;  
 KW immunodominant epitope; human cellular tumour antigen;  
 KW transformation-associated protein.  
 OS Homo sapiens.  
 PN WO9410306-A.  
 PD 11-MAY-1994.  
 PE 02-NOV-1993; F01082.  
 PR 02-NOV-1992; FR-013110.  
 PA (EURO-) LAB EUROBIOS SA.  
 PI Legros Y, Lubin R, Soussi T;  
 DR WPI: 94-167463/20  
 PT New immunodominant epitope(s) of protein p53 - for detecting and  
 PT monitoring antibodies indicative of cancer and precancerous  
 PT states  
 PS Claim 5; Page 42; 62pp; French.  
 CC Peptides derived from the N-terminal (amino acids 1-112) or the C-  
 CC terminal (amino acids 350-393) of protein p53 which specifically  
 CC react with anti-p53 antibodies in patients with cancer or  
 CC precancerous conditions are claimed. The peptides (R54907-R54921)  
 CC are useful for detecting and monitoring cancerous and precancerous  
 CC conditions.  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 67; DB 10; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.22e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 lpennv1spl 15  
 1 lpennv1spl 10  
 QY 1 lpennv1spl 10

RESULT 3  
 ID R54907 standard; peptide; 25 AA.  
 AC R54907;  
 DT 29-NOV-1994 (first entry)  
 DE Immunodominant epitope from p53 N-terminal.  
 KW cancer; pre-cancerous state; detection; diagnosis; human p53 gene;  
 KW immunodominant epitope; human cellular tumour antigen;  
 KW transformation-associated protein.  
 OS Homo sapiens.  
 PN WO9410306-A.  
 PD 11-MAY-1994.  
 PE 02-NOV-1993; F01082.  
 PR 02-NOV-1992; FR-013110.  
 PA (EURO-) LAB EUROBIOS SA.  
 PI Legros Y, Lubin R, Soussi T;  
 DR WPI: 94-167463/20.  
 PT New immunodominant epitope(s) of protein p53 - for detecting and  
 PT monitoring antibodies indicative of cancer and precancerous  
 PT states  
 PS Claim 4; Page 42; 62pp; French.  
 CC Peptides derived from the N-terminal (amino acids 1-112) or the C-  
 CC terminal (amino acids 350-393) of protein p53 which specifically  
 CC react with anti-p53 antibodies in patients with cancer or  
 CC precancerous conditions are claimed. The peptides (R54907-R54921)  
 CC are useful for detecting and monitoring cancerous and precancerous  
 CC conditions.  
 SQ Sequence 25 AA;

Query Match 100.0%; Score 67; DB 10; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 1.22e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 16 lpennv1spl 25  
 1 lpennv1spl 10  
 QY 1 lpennv1spl 10

RESULT 4  
 ID W13604 standard; peptide; 36 AA.  
 AC W13604;  
 DT 16-JAN-1998 (first entry)  
 DE p53 protein amino acids 6-41.  
 KW Mouse; Mdm2; murine double minute; phosphoprotein; binding; modulation;  
 KW tumour suppressor; p53; oncogene; cell cycle arrest; p107; antagonist;  
 KW inhibition; transcription factor; adenocarcinoma; colon; cancer; breast;  
 KW lung; stomach; myeloid leukaemia; lymphoma; hyperproliferative;  
 KW restenosis.  
 OS Homo sapiens.  
 PN WO9709343-A2.  
 PD 13-MAR-1997.  
 PE 02-SEP-1996; F01340.  
 PR 04-SEP-1995; FR-010331.  
 PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 PI (RHON ) RHONE-POULENC RORER SA.  
 DR WPI: 97-192837/17.  
 PT Treating cancer with antagonist of oncogenic activity of protein  
 PT Mdm2 - or nucleic acid encoding an antagonist, also viral vectors  
 PT contg. this nucleic acid  
 PS Claim 4; Page -; 43pp; French.  
 CC The peptides W13602-6 represent peptide fragments derived from the wild  
 CC type human p53 protein. This peptide corresponds to amino acids 6-41  
 CC of the p53 sequence. The peptides are claimed peptides which are able  
 CC to bind the N-terminal amino acids (1-134) of the murine double minute-2  
 CC (mdm2) protein (W13600). Mdm2 protein is a 90 kD phosphoprotein which  
 CC binds and modulates the activity of the tumour suppressor protein p53.  
 CC It has now been shown that the mdm2 protein itself has oncogenic  
 CC properties, especially in a p53-null background. Mdm2 is observed to  
 CC unblock cell cycle arrest in G1 caused by over-expression of the p107  
 CC protein. The p53 peptides are examples of antagonists of the invention  
 CC which are able to inhibit the oncogenic activity of mdm2. The antagonists  
 CC are used to treat e.g. adenocarcinoma of the colon; cancer of the breast,  
 CC lung or stomach; myeloid leukaemia; B cell lymphoma, or other  
 CC hyperproliferative conditions such as restenosis.  
 CC Note: this sequence is not given in the specification but is constructed  
 CC from the wild type human p53 sequence.  
 SQ Sequence 36 AA;

Query Match 100.0%; Score 67; DB 25; Length 36;  
 Best Local Similarity 100.0%; Pred. No. 1.22e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 lpennv1spl 30  
 1 lpennv1spl 10  
 QY 1 lpennv1spl 10

RESULT 5  
 ID W13603 standard; peptide; 41 AA.  
 AC W13603;  
 DT 16-JAN-1998 (first entry)  
 DE p53 protein amino acids 1-41.  
 KW Mouse; Mdm2; murine double minute; phosphoprotein; binding; modulation;  
 KW tumour suppressor; p53; oncogene; cell cycle arrest; p107; antagonist;  
 KW inhibition; transcription factor; adenocarcinoma; colon; cancer; breast;  
 KW lung; stomach; myeloid leukaemia; lymphoma; hyperproliferative;  
 KW restenosis.  
 OS Homo sapiens.  
 PN WO9709343-A2.  
 PD 13-MAR-1997.  
 PE 02-SEP-1996; F01340.  
 PR 04-SEP-1995; FR-010331.  
 PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 PI (RHON ) RHONE-POULENC RORER SA.  
 DR WPI: 97-192837/17.  
 PT Treating cancer with antagonist of oncogenic activity of protein  
 PT Mdm2 - or nucleic acid encoding an antagonist, also viral vectors  
 PT contg. this nucleic acid  
 PS Claim 4; Page -; 43pp; French.

CC The peptides W13602-6 represent peptide fragments derived from the wild  
 CC type human p53 protein. This peptide corresponds to amino acids 1-41  
 CC of the p53 sequence. The peptides are claimed peptides which are able  
 CC to bind the N-terminal amino acids (1-134) of the murine double minute-2  
 CC (mdm2) protein (W13600). Mdm2 protein is a 90 kD phosphoprotein which  
 CC binds and modulates the activity of the tumour suppressor protein p53.  
 CC It has now been shown that the mdm2 protein itself has oncogenic  
 CC properties, especially in a p53-null background. Mdm2 is observed to  
 CC unblock cell cycle arrest in G1 caused by over-expression of the p107  
 CC protein. The p53 peptides are examples of antagonists of the invention  
 CC which are able to inhibit the oncogenic activity of mdm2. The antagonists  
 CC are used to treat e.g. adenocarcinoma of the colon; cancer of the breast,  
 CC lung or stomach; myeloid leukaemia; B cell lymphoma, or other  
 CC hyperproliferative conditions such as restenosis.  
 CC Note: this sequence is not given in the specification but is constructed  
 CC from the wild type human p53 sequence.  
 SO Sequence 41 AA;

Query Match 100.0%; Score 67; DB 25; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 1.22e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 26 lpenntlsp1 35  
 |||||||||  
 QY 1 LPENNTLSPL 10

RESULT 6  
 ID W13602 standard; peptide: 52 AA.  
 AC W13602;  
 DE 16-JAN-1998 (first entry)  
 DE p53 protein amino acids 1-52.  
 KW Mouse; Mdm2; murine double minute; phosphoprotein; binding; modulation;  
 KW tumour suppressor; p53; oncogene; cell cycle arrest; p107; antagonist;  
 KW inhibition; transcription factor; adenocarcinoma; colon; cancer; breast;  
 KW lung; stomach; myeloid leukaemia; lymphoma; hyperproliferative;  
 KW restenosis.  
 OS Homo sapiens.  
 PN MO9709343-A2.  
 PD 13-MAR-1997.  
 PE 02-SEP-1996; F01340.  
 PR 04-SEP-1995; FR-010331.  
 PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 PA (RHON ) RHONE-POULENC RORER SA.  
 PI Jubs-Poterszman M, Tocque B, Wasylyk B;  
 DR WPI: 97-192837/17.  
 PT Treating cancer with antagonist of oncogenic activity of protein  
 PT Mdm2 - or nucleic acid encoding an antagonist, also viral vectors  
 PT conty. this nucleic acid  
 PS Claim 4; Page -; 43pp; French.  
 CC The peptides W13602-6 represent peptide fragments derived from the wild  
 CC type human p53 protein. This peptide corresponds to amino acids 1-52  
 CC of the p53 sequence. The peptides are claimed peptides which are able  
 CC to bind the N-terminal amino acids (1-134) of the murine double minute-2  
 CC (mdm2) protein (W13600). Mdm2 protein is a 90 kD phosphoprotein which  
 CC binds and modulates the activity of the tumour suppressor protein p53.  
 CC It has now been shown that the mdm2 protein itself has oncogenic  
 CC properties, especially in a p53-null background. Mdm2 is observed to  
 CC unblock cell cycle arrest in G1 caused by over-expression of the p107  
 CC protein. The p53 peptides are examples of antagonists of the invention  
 CC which are able to inhibit the oncogenic activity of mdm2. The antagonists  
 CC are used to treat e.g. adenocarcinoma of the colon; cancer of the breast,  
 CC lung or stomach; myeloid leukaemia; B cell lymphoma, or other  
 CC hyperproliferative conditions such as restenosis.  
 CC Note: this sequence is not given in the specification but is constructed  
 CC from the wild type human p53 sequence.  
 SO Sequence 52 AA;

Query Match 100.0%; Score 67; DB 25; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.22e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 26 lpenntlsp1 35

QY |||||||||  
 1 LPENNTLSPL 10

RESULT 7  
 ID W42878 standard; Protein: 64 AA.  
 AC W42878;  
 DE 30-APR-1998 (first entry)  
 DE N-terminal region of p53 (amino acids 13-41).  
 KW MDM2; tumour; diagnosis; neoplasia; DNA binding protein;  
 KW p53 polypeptide; binding; tumour cell; p53-regulated growth;  
 KW inhibition; anti-cancer agent.  
 OS Homo sapiens.  
 PN US5708136-A.  
 PD 13-JAN-1998.  
 PE 17-FEB-1995; 390516.  
 PR 07-APR-1993; US-044619.  
 PR 07-APR-1992; US-867840.  
 PR 23-JUN-1992; US-903103.  
 PA (UYXO ) UNIV JOHNS HOPKINS.  
 PI Burrell M, Hill DE, Kinzler KW, Vogelstein B;  
 DR WPI: 98-100408/09.  
 PT Human MDM2 binding polypeptide - comprises fragments of p53, useful  
 PT in re-establishing p53-regulated growth control in cells  
 PT over-expressing MDM2  
 PS Claim 1; Columns 19-20; 41pp; English.  
 CC The present sequence represents a N-terminal portion of the p53 protein.  
 CC These amino acid residues have been found to be necessary for the  
 CC interaction of MDM2 and p53. A cell containing three recombinant  
 CC DNA constructs was produced. These constructs encode an MDM2 protein  
 CC fused to a sequence-specific DNA binding domain, a p53 polypeptide fused  
 CC to a transcriptional activation domain, and a reporter gene downstream  
 CC from a DNA element which is recognised by the sequence-specific  
 CC DNA-binding domain. The cell is used to identify a compound which  
 CC interferes with the binding of MDM2 and p53. Since MDM2 is overexpressed  
 CC in tumour cells and since binding of MDM2 to p53 appears to allow tumour  
 CC cells to escape from p53-regulated growth, compounds that inhibit such  
 CC binding would be useful as anti-cancer agents.  
 SO Sequence 64 AA;

Query Match 100.0%; Score 67; DB 28; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 1.22e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 26 lpenntlsp1 35  
 |||||||||  
 QY 1 LPENNTLSPL 10

RESULT 8  
 ID W48243 standard; peptide: 64 AA.  
 AC W48243;  
 DE 18-JUN-1998 (first entry)  
 DE Human p53 fragment 13-41.  
 KW Human; MDM2; hMDM2; tumour; cancer; diagnosis; neoplastic disease; p53;  
 KW sarcoma; liposarcoma; malignant fibrous histiocytoma; osteosarcoma.  
 OS Homo sapiens.  
 PN US5736338-A.  
 PD 07-APR-1998.  
 PE 17-FEB-1995; 390517.  
 PR 07-APR-1993; US-044619.  
 PR 07-APR-1992; US-867840.  
 PR 23-JUN-1992; US-903103.  
 PR 17-FEB-1995; US-390517.  
 PA (UYXO ) UNIV JOHNS HOPKINS.  
 PI Burrell M, Hill DE, Kinzler KW, Vogelstein B;  
 DR WPI: 98-229206/21.  
 PT Cancer diagnosis - by determination of MDM2 protein  
 PS Disclosure; Column 19-20; 35pp; English.  
 CC The present sequence represents a human p53 fragment 13-41 which is  
 CC necessary for the interaction of MDM2 and p53. The present invention  
 CC describes a method for diagnosing a neoplastic disease caused by  
 CC overexpression of MDM2 protein. The method comprises detecting an

CC elevated cellular amount of this protein. The method is useful for the  
 CC diagnosis of sarcoma, especially liposarcoma, malignant fibrous  
 CC histiocytoma or osteosarcoma.  
 SO Sequence 64 AA.

Query Match 100.0%; Score 67; DB 29; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 1.22e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 lpennvlspl 35  
 |||||||  
 QY 1 LPENNVLSPL 10

RESULT 9  
 ID W42970 standard; Protein: 64 AA.  
 AC W42970;  
 DT 29-APR-1998 (first entry)  
 DE N-terminal region of p53 (amino acids 13-41).  
 KM MDM2; tumour; diagnosis; neoplasia; DNA binding protein;  
 KW p53 polypeptide; binding; tumour cell; p53-regulated growth;  
 CC inhibition; anti-cancer agent.  
 OS Homo sapiens.  
 PN US5702903-A.  
 PD 30-DEC-1997.  
 PE 13-NOV-1995; 557393.  
 PR 07-APR-1993; US-044619.  
 PR 07-APR-1992; US-867840.  
 PR 23-JUN-1992; US-903103.  
 PR 18-MAY-1994; US-245500.  
 PA (UJJO ) UNIV JOHNS HOPKINS.  
 PI Kinzler KW, Vogelstein B;  
 DR WPI: 98-076411/07.

PT Cell containing reporter construct containing human MDM2 and p53  
 PT genes - for identifying compounds that interfere with binding of  
 PT human MDM2 to human p53, useful as anti-cancer agents  
 PS Disclosure: Coulms 19-20; 37pp; English.  
 CC The present sequence represents a N-terminal portion of the p53 protein.  
 CC These amino acid residues have been found to be necessary for the  
 CC interaction of MDM2 and p53. A cell containing three recombinant  
 CC DNA constructs was produced. These constructs encode an MDM2 protein  
 CC fused to a sequence-specific DNA binding domain, a p53 polypeptide fused  
 CC to a transcriptional activation domain, and a reporter gene downstream  
 CC from a DNA element which is recognised by the sequence-specific  
 CC DNA-binding domain. The cell is used to identify a compound which  
 CC interferes with the binding of MDM2 and p53. Since MDM2 is overexpressed  
 CC in tumour cells and since binding of MDM2 to p53 appears to allow tumour  
 CC cells to escape from p53-regulated growth, compounds that inhibit such  
 CC binding would be useful as anti-cancer agents.  
 SO Sequence 64 AA;

Query Match 100.0%; Score 67; DB 28; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 1.22e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 lpennvlspl 35  
 |||||||  
 QY 1 LPENNVLSPL 10

RESULT 10  
 ID W07886 standard; Protein: 64 AA.  
 AC W07886;  
 DT 28-JAN-1997 (first entry)  
 DE Human p53; involved in tumour suppression.  
 KM p53; MDM-2; binding-inhibitor; identification; tumour; cancer;  
 KW neoplasia; antibody fusion protein; therapy.  
 OS Homo sapiens.  
 FH Key  
 FT region  
 FT region  
 FT region  
 FT region

Location/Qualifiers  
 1..41  
 /note- "MDM-2 binding fragment"  
 1..50  
 /note- "MDM-2 binding fragment"

FT region 13..57  
 /note- "MDM-2 binding fragment"

PN US5550023-A.  
 PD 27-AUG-1996.  
 PE 07-APR-1992; 867840.  
 PR 07-APR-1992; US-867840.  
 PR 23-JUN-1992; US-903103.  
 PR 07-APR-1993; US-044619.  
 PR 18-MAY-1994; US-245500.  
 PA (UJJO ) UNIV JOHNS HOPKINS.  
 PI Kinzler KW, Vogelstein B;  
 DR WPI: 96-401591/40.  
 PT Identification of cpds. interfering with human MDM2/p53 binding -  
 PT useful as therapeutic agents to treat human neoplastic cells  
 PS Claim 16. Column 19-20; 36pp; English.  
 CC W07886 represents the human p53 protein which is involved in the  
 CC development of many cancers. The protein is used here in a method  
 CC for identifying compounds that interfere with the binding of p53 and  
 CC MDM-2. In binding the p53 protein, the MDM-2 protein releases a cell  
 CC from p53-regulated growth, allowing cancers to develop. Therefore  
 CC compounds identified as interfering with the binding of MDM-2 to p53  
 CC are potentially useful in the treatment of human neoplastic cells. In  
 CC the method pref. one or both of the proteins is a fusion protein esp.  
 CC with an antibody or antibody fragment which aids separation and  
 CC identification.  
 SO Sequence 64 AA;

Query Match 100.0%; Score 67; DB 19; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 1.22e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 lpennvlspl 35  
 |||||||  
 QY 1 LPENNVLSPL 10

RESULT 11  
 ID W47079 standard; peptide: 71 AA.  
 AC W47079;  
 DT 19-MAY-1998 (first entry)  
 DE Human p53 fragment (residues 1-71)  
 KM Retinoblastoma gene; RB; p53 protein; interaction; inhibitor;  
 KW tumour; apoptosis; modulator; medicine; veterinary; human.  
 OS Homo sapiens.  
 PN M09741433-A1.  
 PD 06-NOV-1997.  
 PE 29-APR-1997; G01168.  
 PE 29-APR-1996; GB-008937.  
 PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.  
 PI Kouzarides T;  
 DR WPI: 97-549887/50.

PT Identifying compounds that modulate interaction of p53 and RB  
 PT protein - or those that bind to RB protein, used to induce  
 PT apoptosis, specifically for treatment of tumours  
 PS Disclosure: Fig 3B; 83pp; English.

CC This is a fragment of the human p53 protein. p53 peptide fragments  
 CC acting as inhibitors of the interaction between a p53 protein and a  
 CC retinoblastoma (RB) protein are synthesised based on this sequence. The  
 CC interaction between p53 and RB is found to be critical for determining  
 CC whether or not a cell enters apoptosis. Apoptosis is prevented if  
 CC interaction occurs. The interaction is between regions 1-71 or 290-393  
 CC of p53 and region 379-928 of RB. The invention provides methods to  
 CC identify compounds able to modulate interaction or binding between p53  
 CC and RB protein. The method comprises combining p53 and RB, or their  
 CC fragments, with a test compound and detecting interaction/binding between  
 CC them. The inhibitory compounds are used in human or veterinary medicine  
 CC to modulate p53 activity and processes, specifically for inducing  
 CC apoptosis in tumour cells (possibly also in cells infected by virus),  
 CC in vivo or in vitro. Expression of these modulators by gene therapy  
 CC methods is also contemplated. Other activities that can be affected are  
 CC transcription repression, G1 arrest, DNA repair, homologous recombination  
 CC and 3'-5'-exonuclease activity. Modulation of interaction with RB may  
 CC also stabilise p53.

|        |  |
|--------|--|
| RESULT | 13   |
| ID     | W13950 standard; Protein; 355 AA.  |
| AC     | W13950;  |
| DT     | 25-JUN-1997 (first entry)  |
| DE     | Del1356-393 modified human p53.  |
| KW     | p53; tumour suppressor; cancer; therapy; cell proliferation;                           |
| KW     | apoptosis; protein engineering; DNA binding.   |
| OS     | Synthetic  |
| PN     | MO9710683-A1.  |
| PD     | 27-MAR-1997.   |
| PF     | 20-SEP-1996; U15188.   |
| PR     | 22-SEP-1995; US-004802.  |
| PR     | 21-AUG-1996; US-697221.  |
| PA     | (WIST-) WISTAR INST ANATOMY & BIOLOGY.   |
| PI     | Halazonecis TD;  |
| PT     | R284K modified p53. protein having DNA binding ability - useful in treatment of cancer |
| PS     | Claim 3: Refer to page 27-29; 82pp; English.   |
| CC     | Del1356-393 modified p53 (W13950) has the C-terminal region of                         |
| CC     | wild-type human p53 tumour suppressor (W13948) deleted. Modified                       |
| CC     | p53 constructs (see also W13954, W13956-61, W13971-77) bearing                         |
| CC     | a deletion of all or a fragment of the C-terminal residues                             |
| CC     | 356-393 have DNA binding ability and can activate the DNA binding                      |
| CC     | of common Class I p53 tumour mutants (see also W13951-52). The                         |

|        |  |
|--------|--|
| RESULT | 15   |
| ID     | w13961 standard; protein; 361 AA.                          |
| AC     | w13961;  |
| DT     | 25-JUN-1997 (first entry)                                  |
| DE     | Chimeric p53 protein.                                      |
| KW     | p53; tumor suppressor; cancer; therapy; cell proliferation |
| KM     | apoptosis; protein engineering; GCNA; DNA binding.         |
| OS     | Chimeric Homo sapiens;                                     |
| OS     | Chimeric synthetic.  |
| RH     | key Location/Qualifiers                                    |

```

FT      region      1..323
FT      /label= p53wt
FT      /note= "amino acids 1-323 of wild-type p53"
FT      region      324..329
FT      /label= linker
FT      region      330..361
FT      /label= GCN4
FT      /note= "amino acids 250-281 of GCN4 LZ variant"
PN      MO9710843-A1
PD      27-MAR-1997
PF      20-SEP-1986: U15188.
PA      22-SEP-1995: US-004802.
PR      21-AUG-1996: US-697221.
PI      (WIST-) WISTAR INSTR ANATOMY & BIOLOGY.
PI      Halazoneitis TD;
DR      WPI: 97-202618/18.
PT      R284K modified p53 protein having DNA binding ability - useful in
PT      treatment of cancer
PS      Disclosure: Refer to Page 8; 82pp; English.
CC      Chimeric p53 constructs (W13956-67) comprise N-terminal portions
CC      of human wild-type p53 tumour suppressor (see also W13948) linked
CC      to a C-terminal portion of the LZ variant (see also W13955) of
CC      GCN4 and, in some cases, the C-terminal portion of wild-type
CC      p53. The chimeric proteins have DNA binding activity and can
CC      replace lost or insufficient p53 function, providing the means for
CC      pharmacological rescue of p53 function in cancer patients. Nucleic
CC      acids coding for modified p53 constructs can be used for cancer
CC      gene therapy.
SQ      Sequence 361 AA:

```

```

Query Match      100.0%; Score 67; DB 21; Length 361;
Best Local Similarity 100.0%; Pred. No. 1.22e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db: 26 lpennvlspl 35
    |||||
QY 3 LPENNVLSP 10

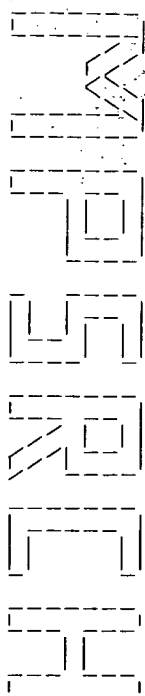
```

```

Search completed: Fri Sep 11 13:25:08 1998
Job time: 15 secs.

```





(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch.p protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Sep 11 13:25:26 1998; Maspar time 3.06 Seconds  
Tabular output not generated. 119.480 Million cell updates/sec

Title: >US-08-452-843-13  
Description: (1-10) from US08452843.pep  
Perfect Score: 67  
Sequence: 1 LPENNVLSPL 10

Scoring table: PAM 150  
Gap 15

Searched: 120441 segs, 36531193 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: p1r56  
1:pir1 2:pir2 3:pir3 4:pir4 5:nrl3d

Statistics: Mean 22.847; Variance 30.062; scale 0.760

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description            | Pred. No. |
|------------|-------|-------------|--------|----------|------------------------|-----------|
| 1          | 67    | 100.0       | 393    | 1 DNHU53 | cellular tumor antigen | 1.97e+03  |
| 2          | 67    | 100.0       | 393    | 2 S06594 | cellular tumor antigen | 1.97e+03  |
| 3          | 51    | 76.1        | 393    | 2 JC6176 | tumor suppressor prot  | 5.01e+00  |
| 4          | 51    | 76.1        | 396    | 2 JH0633 | cellular tumor antigen | 5.01e+00  |
| 5          | 51    | 76.1        | 1752   | 2 S26849 | DNA-directed RNA poly  | 5.01e+00  |
| 6          | 50    | 74.6        | 785    | 2 S03785 | dimethylsulfoxide red  | 7.85e+00  |
| 7          | 49    | 73.1        | 1294   | 2 S77690 | probable membrane pro  | 1.22e+01  |
| 8          | 48    | 71.6        | 162    | 2 S49775 | hypothetical protein   | 1.89e+01  |
| 9          | 48    | 71.6        | 409    | 2 S52095 | tau-protein kinase (E  | 1.89e+01  |
| 10         | 48    | 71.6        | 527    | 2 F69378 | conserved hypothetical | 1.89e+01  |
| 11         | 47    | 70.1        | 640    | 2 S75175 | hypothetical protein   | 2.91e+01  |
| 12         | 47    | 70.1        | 101    | 2 S45299 | hypothetical protein   | 2.91e+01  |
| 13         | 47    | 70.1        | 207    | 2 JC4042 | lexa protein - Aetomo  | 2.91e+01  |
| 14         | 47    | 70.1        | 320    | 2 B69272 | conserved hypothetical | 2.91e+01  |
| 15         | 47    | 70.1        | 394    | 2 D69370 | acyl-CoA dehydrogenas  | 2.91e+01  |
| 16         | 47    | 70.1        | 577    | 2 F65202 | hypothetical 66.6 kD   | 2.91e+01  |
| 17         | 47    | 70.1        | 1124   | 2 B45557 | pol polyprotein - fel  | 2.91e+01  |
| 18         | 46    | 68.7        | 112    | 2 S24417 | ferredoxin-like prot   | 4.45e+01  |
| 19         | 46    | 68.7        | 198    | 2 S33928 | early E1A protein - h  | 4.45e+01  |
| 20         | 46    | 68.7        | 266    | 1 A0AD62 | reverse transcriptase  | 4.45e+01  |
| 21         | 46    | 68.7        | 318    | 2 S58503 | cellular tumor antigen | 4.45e+01  |
| 22         | 46    | 68.7        | 386    | 2 S51648 | complement C3b/C4b re  | 4.45e+01  |
| 23         | 46    | 68.7        | 482    | 2 A34924 |                        |           |

|    |    |      |      |          |                       |          |
|----|----|------|------|----------|-----------------------|----------|
| 24 | 46 | 68.7 | 537  | 2 B46535 | interleukin 2 recepto | 4.45e+01 |
| 25 | 45 | 67.2 | 160  | 2 S44736 | b0523.2 protein - Cae | 6.75e+01 |
| 26 | 45 | 67.2 | 348  | 2 S44628 | f22b7.1 protein - Cae | 6.75e+01 |
| 27 | 45 | 67.2 | 640  | 1 QYCHGM | phosphoenolpyruvate c | 6.75e+01 |
| 28 | 45 | 67.2 | 860  | 2 S55543 | reverse transcriptase | 6.75e+01 |
| 29 | 45 | 67.2 | 908  | 2 S07649 | gene coi intron 1 pro | 6.75e+01 |
| 30 | 45 | 67.2 | 1124 | 2 S23820 | pol polyprotein - fel | 6.75e+01 |
| 31 | 45 | 67.2 | 1124 | 1 GNLJRP | pol polyprotein - fel | 6.75e+01 |
| 32 | 45 | 67.2 | 1252 | 2 S77037 | hypothetical protein  | 6.75e+01 |
| 33 | 44 | 65.7 | 262  | 2 S60213 | homc protein - Strept | 1.02e+02 |
| 34 | 44 | 65.7 | 303  | 2 S28392 | protein-tyrosine-phos | 1.02e+02 |
| 35 | 44 | 65.7 | 304  | 1 CRVZW  | cell surface-binding  | 1.02e+02 |
| 36 | 44 | 65.7 | 304  | 1 CRVZ7P | cell surface-binding  | 1.02e+02 |
| 37 | 44 | 65.7 | 317  | 2 A37388 | probable DNA-binding  | 1.02e+02 |
| 38 | 44 | 65.7 | 317  | 2 B37388 | hypothetical protein  | 1.02e+02 |
| 39 | 44 | 65.7 | 388  | 2 E65054 | hypothetical protein  | 1.02e+02 |
| 40 | 44 | 65.7 | 452  | 2 S77538 | serine proteinase (EC | 1.02e+02 |
| 41 | 44 | 65.7 | 468  | 2 A55476 | protein kinase (EC 2. | 1.02e+02 |
| 42 | 44 | 65.7 | 502  | 2 S26004 | 18S rRNA intron 1 pro | 1.02e+02 |
| 43 | 44 | 65.7 | 793  | 1 A25691 | glucocorticoid recept | 1.02e+02 |
| 44 | 44 | 65.7 | 795  | 1 QRRTE  | glucocorticoid recept | 1.02e+02 |
| 45 | 44 | 65.7 | 806  | 2 G64109 | dimethylsulfoxide red | 1.02e+02 |

## ALIGNMENTS

| RESULT            | 1  | ALIGNMENTS     |
|-------------------|--|----------------|
| ENTRY             | DNH53  | #type complete |
| TITLE             | cellular tumor antigen p53 - human   |                |
| ALTERNATE_NAMES   | cellular phosphoprotein p53; oncoprotein p53; transformation suppressor p53; tumor suppressor p53  |                |
| ORGANISM          | #formal_name Homo sapiens #common_name man   |                |
| DATE              | 05-Oct-1988 #sequence_revision 18-Nov-1994 #text_change 18-Sep-1997  |                |
| ACCESSIONS        | A25224; A43073; J0436; S40773; S42669; A22837; A5060; A25397; B25397; S42452; S42452; I38082; I38083; I38084; I38085; I38086; I38087; I38088; I38089; I38090; I38091; I38092; I38093; A44905; I58354; I78850; S60153 |                |
| REFERENCE         | A25224   |                |
| #authors          | Lamb, P.; Crawford, L.   |                |
| #journal          | Mol. Cell. Biol. (1986) 6:1379-1385  |                |
| #title            | Characterization of the human p53 gene.  |                |
| #cross-references | EMBL:87064416  |                |
| #accession        | A25224   |                |
| #molecule_type    | DNA  |                |
| #residues         | 1-393 ##label LAM  |                |
| #cross-references | EMBL:X01405; GB:M13121; GB:N00032; NID:g189460; PDB:g386994  |                |
| REFERENCE         | J0436  |                |
| #authors          | Buchanan, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.; Georgiev, G.P.  |                |
| #journal          | Gene (1988) 70:245-252   |                |
| #title            | A variation in the structure of the protein-coding region of the human p53 gene.   |                |
| #cross-references | EMBL:89108008  |                |
| #accession        | A43073   |                |
| #molecule_type    | DNA  |                |
| #residues         | 1-393 ##label BUC  |                |
| #note             | this 72-Arg allele appears to be about 5 times more frequent than the 72-Pro allele  |                |
| #accession        | J0436  |                |
| #molecule_type    | DNA  |                |
| #residues         | 1-71, 'p', '73-393 ##label B02   |                |
| #cross-references | EMBL:M22898; NID:g189474; PDB:g189476  |                |
| #note             | this 72-Pro allele was found in both normal and malignant cell lines   |                |
| REFERENCE         | S40773   |                |
| #authors          | Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.   |                |
| #submission       | submitted to the EMBL Data Library, August 1990  |                |
| #accession        | S40773   |                |
| #molecule_type    | DNA  |                |
| #residues         | 1-393 ##label CHU  |                |
| #cross-references | EMBL:X54156; NID:g35213; PDB:g35214  |                |

REFERENCE S42669  
#authors Metlshewski, G.; Lamb, P.; Pim, D.; Peacock, J.; Crawford, L.; Benchimol, S.  
#journal EMBO J. (1984) 3:3257-3262  
#title Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene.  
#accession S42669  
##molecule-type mRNA  
##residues 101-393 ##label MK1  
##cross-references EMBL:X01405; NID:935215; PID:9642241  
REFERENCE A22837  
#authors Zakut-Hourli, R.; Bienz-Tadmor, B.; Girol, D.; Oren, M.  
#journal EMBO J. (1985) 4:1251-1255  
#title Human p53 cellular tumor antigen: cDNA sequence and expression in COS cells.  
#cross-references M01D:8530577  
#accession A22837  
##molecule-type mRNA  
##residues 1-71, 'P', '73-393 ##label ZAK  
##cross-references EMBL:X02469; EMBL:M60950; NID:935209; PID:935210  
REFERENCE A55060  
#authors Harlow, E.; Williamson, N.M.; Ralston, R.; Hellman, D.M.; Adams, T.E.  
#journal Mol. Cell. Biol. (1985) 5:1601-1610  
#title Molecular cloning and in vitro expression of a cDNA clone for human cellular tumor antigen p53.  
#accession A55060  
##molecule-type mRNA  
##residues 1-71, 'P', '73-272, 'H', 274-393 ##label HA3  
##cross-references GB:K03199; NID:9189478; PID:9189479  
##experimental-source clone pR4-2, cell line A431  
REFERENCE A93086  
#authors Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.; Arad, N.; Rotter, Y.  
#journal Mol. Cell. Biol. (1986) 6:4650-4656  
#title Molecular basis for heterogeneity of the human p53 protein.  
#cross-references M01D:87089826  
#accession A25397  
##molecule-type mRNA  
##residues 1-78, 'T', 80-393 ##label HAR  
##cross-references EMBL:M14694; NID:9339813; PID:9339814  
##experimental-source clone p53-H-1, transformed hybridoma SV-80 cell line  
#accession B25397  
##molecule-type mRNA  
##residues 1-71, 'P', '73-78, 'T', 80-393 ##label HA2  
##cross-references EMBL:M14695; NID:9339815; PID:9339816  
##experimental-source clone p53-H-19, transformed hybridoma SV-80 cell line  
REFERENCE S42452  
#authors Metlshewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider, J.; Crawford, L.V.  
#journal Mol. Cell. Biol. (1987) 7:961-963  
#title Primary structure polymorphism at amino acid residue 72 of human p53.  
#accession S42452  
##molecule-type DNA  
##residues 66-71, 'P', '73-79 ##label MK2  
##experimental-source clone lambda C113  
##note 72-Cys was also found, and appears to represent a polymorphism  
#accession S42453  
##molecule-type mRNA: DNA  
##residues 66-79 ##label MAT  
##experimental-source clone J6K  
REFERENCE I38082  
#authors Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.; Crook, T.  
#journal EMBO J. (1991) 10:2879-2887  
#title p53 is frequently mutated in Burkitt's lymphoma cell lines.  
#cross-references M01D:92007731  
#accession I38082  
#status translated from GB/EMBL/DBJ

##molecule-type mRNA  
##residues 1-189, 'L', '189-566  
##cross-references EMBL:X60010; NID:9506432; PID:9506433  
##note deletion of a C nucleotide causes a frameshift at position 566  
#accession I38083  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-192, 'R', 194-393 ##label F02  
##cross-references EMBL:X60011; NID:9506434; PID:9506435  
#accession I38084  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-393 ##label F03  
##cross-references EMBL:X60012; NID:9506436; PID:9506437  
#accession I38085  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-245, 'T', 247-393 ##label F04  
##cross-references EMBL:X60013; NID:9506438; PID:9506439  
#accession I38086  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-236, 'I', 238-393 ##label F05  
##cross-references EMBL:X60014; NID:9506440; PID:9506441  
#accession I38087  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-247, 'Q', 249-393 ##label F06  
##cross-references EMBL:X60015; NID:9506442; PID:9506443  
#accession I38088  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-71, 'P', '73-237, 'Y', 239-393 ##label F07  
##cross-references EMBL:X60016; NID:9506444; PID:9506445  
#accession I38089  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-247, 'Q', 249-393 ##label F08  
##cross-references EMBL:X60017; NID:9506446; PID:9506447  
#accession I38090  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-71, 'P', '73-162, 'H', 164-393 ##label F09  
##cross-references EMBL:X60018; NID:9506448; PID:9506449  
#accession I38091  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-212, 'Q', 214-393 ##label F10  
##cross-references EMBL:X60019; NID:9506450; PID:9506451  
#accession I38092  
##status translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-253, 'D', 255-393 ##label F11  
##cross-references EMBL:X60020; NID:9506452; PID:9506453  
##note all sequences submitted to the EMBL/GenBank/DBJ databases June 1991  
REFERENCE I38093  
#authors Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.  
#journal Nucleic Acids Res. (1991) 19:6977  
#title An Alu polymorphism intragenic to the TP53 gene.  
#cross-references M01D:92107726  
#accession I38093  
##status translated from GB/EMBL/DBJ  
##molecule-type DNA  
##residues 1-393 ##label RE2  
##cross-references EMBL:X54156; NID:935213; PID:935214  
REFERENCE A44905  
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.; Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.; Terada, M.  
#journal Cancer Res. (1991) 51:5800-5805

#title p53 gene mutations in gastric cancer metastases and in gastric cancer cell lines derived from metastases.  
 #cross-references MUID:92034678  
 #accession M44905  
 #molecule\_type DNA  
 #residues 246-247, 'W', 249-250 #label YAM  
 #cross-references GB:S63157; NID:9237829; PID:9237830  
 #note sequence extracted from NCBI backbone (NCBIN:63157,

Note: remainder of annotations omitted.

Query Match ... 100.0%; Score 67; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1.97e-03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
 |||||||  
 QY 1 LPENNVSPL 10

RESULT 2  
 ENTRY S06594 #type complete  
 TITLE cellular tumor antigen p53 - green monkey  
 ORGANISM #formal\_name Cercopithecus aethiops #common\_name green monkey, grivet  
 DATE 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 08-Sep-1997

ACCESSIONS S06594  
 REFERENCE S06594  
 #authors Rigaudy, P.; Eckhart, W.  
 #journal Nucleic Acids Res. (1989) 17:8375  
 #title Nucleotide sequence of a cDNA encoding the monkey cellular phosphoprotein p53.  
 #cross-references MUID:90045967  
 #accession S06594  
 #molecule\_type mRNA  
 #residues 1-393 #label RIG  
 ##cross-references EMBL:X16384; NID:922795; PID:922796  
 CLASSIFICATION #superfamily cellular tumor antigen p53  
 KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; nucleus; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 176,179,238,242 #binding\_site zinc (Cys, His, Cys, Cys) #status predicted

392 #binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted  
 SUMMARY #length 393 #molecular\_weight 43696 #checksum 4263

Query Match 100.0%; Score 67; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1.97e-03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
 |||||||  
 QY 1 LPENNVSPL 10

RESULT 3  
 ENTRY JC6176 #type complete  
 TITLE tumor suppressor protein p53 - Chinese hamster  
 ORGANISM #formal\_name Cricetus griseus #common\_name Chinese hamster  
 DATE 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 08-Sep-1997

ACCESSIONS JC6176  
 REFERENCE JC6176  
 #authors Lee, H.; Larner, J.M.; Hamlin, J.L.  
 #journal Gene (1997) 184:177-183  
 #title Cloning and characterization of Chinese hamster p53 cDNA.  
 #contents Liver  
 #accession JC6176  
 #molecule\_type mRNA  
 #residues 1-393 #label LEE

#cross-references GB:U50395; NID:91842229; PID:91842230  
 COMMENT This protein is a multimer, it plays the central role in a complex DNA damage-sensing network. It binds to replication factor and TATA-binding protein, and affects DNA replication, transcription, and recombination by protein/protein interactions.

GENETICS p53  
 CLASSIFICATION #superfamily cellular tumor antigen p53  
 KEYWORDS liver; tumor  
 SUMMARY #length 393 #molecular\_weight 43362 #checksum 4043

Query Match 76.1%; Score 51; DB 2; Length 393;  
 Best Local Similarity 80.0%; Pred. No. 5.01e-00;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
 |||||||  
 QY 1 LPENNVSPL 10

RESULT 4  
 ENTRY JH0633 #type complete  
 TITLE cellular tumor antigen p53 - golden hamster  
 ALTERNATE\_NAMES tumor-suppressor protein p53  
 ORGANISM #formal\_name Mesocricetus auratus #common\_name golden hamster  
 DATE 17-Aug-1992 #sequence\_revision 17-Aug-1992 #text\_change 08-Sep-1997

ACCESSIONS JH0633  
 REFERENCE JH0633  
 #authors Legros, Y.; McIntyre, P.; Soussi, T.  
 #journal Gene (1992) 112:247-250  
 #title The cDNA cloning and immunological characterization of hamster p53.  
 #cross-references MUID:92210007  
 #accession JH0633  
 #molecule\_type mRNA  
 #residues 1-396 #label LEE  
 ##cross-references GB:M75144; NID:9191414; PID:9191415  
 ##experimental\_source kidney, strain MPI

GENETICS p53

CLASSIFICATION #superfamily cellular tumor antigen p53  
 KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; nucleus; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 179,182,241,245 #binding\_site zinc (Cys, His, Cys, Cys) #status predicted  
 395 #binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted

SUMMARY #length 396 #molecular\_weight 43631 #checksum 6617  
 Query Match 76.1%; Score 51; DB 2; Length 396;  
 Best Local Similarity 80.0%; Pred. No. 5.01e-00;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
 |||||||  
 QY 1 LPENNVSPL 10

RESULT 5  
 ENTRY S26849 #type complete  
 TITLE DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - fission yeast (Schizosaccharomyces pombe)  
 ORGANISM #formal\_name Schizosaccharomyces pombe  
 DATE 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 12-Sep-1997

ACCESSIONS S26849  
 REFERENCE S26849  
 #authors Azuma, Y.; Yamagishi, M.; Ueshima, R.; Ishihama, A.  
 #journal Nucleic Acids Res. (1991) 19:461-468  
 #title Cloning and sequence determination of the Schizosaccharomyces

[illegible]

```

Best Local Similarity 60.0%:  Pred No. 7,85e+00:
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 637 LPEGDVIDPL 646
||| :|: ||
Oy 1 LPENNVLSP 110

RESULT 7
ENTRY S77690 #type complete
TITLE probable membrane protein YOL075c - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein O1125; hypothetical protein O1130;
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 20-Feb-1998
ACCESSIONS S77690; S66767; S66768
REFERENCE #authors Alexandrakl, D.; Katsoulou, C.; Tzeremia, M.
#submission submitted to the Protein Sequence Database, July 1996
#accession S77690
##molecule_type DNA
##residues 1-1294 #label ALP
##cros_references EMBL:Z74816; MIPS:YOL075c
#note This is a revision to the sequence from reference S66756
REFERENCE S66756
#authors Alexandrakl, D.; Katsoulou, C.; Tzeremia, M.
#submission submitted to the Protein Sequence Database, July 1996
#accession S66767
##molecule_type DNA
##residues 1-179, 'TTRNGVFLVVKRPD' #label ALW
##cros_references EMBL:Z74816
#experimental_source strain S288C
#note this sequence has been revised in reference S77690
#accession S66768
##molecule_type DNA
##residues 200-1294 #label ALP
##cros_references EMBL:Z74817
#experimental_source strain S288C
#note this sequence has been revised in reference S77690
#note this was assumed to be the complete sequence of protein YOL075c

GENETICS #map_position 15L
#note YOL075c

CLASSIFICATION #superfamily unassigned ATP-binding cassette proteins:
ATP-binding cassette homology
transmembrane protein

KEYWORDS
FEATURE
45-263 #domain ATP-binding cassette homology #label ABC1\
376-392 #domain transmembrane #status predicted #label TM1\
469-485 #domain transmembrane #status predicted #label TM2\
496-512 #domain transmembrane #status predicted #label TM3\
606-622 #domain transmembrane #status predicted #label TM4\
710-916 #domain ATP-binding cassette homology #label ABC2\
1042-1058 #domain transmembrane #status predicted #label TM5\
1125-1141 #domain transmembrane #status predicted #label TM6\
1177-1193 #domain transmembrane #status predicted #label TM7\
1269-1285 #domain transmembrane #status predicted #label TM8\
SUMMARY #length 1294 #molecular_weight 145156 #checksum 3044

Query Match 73.1%: Score 49; DB 2; Length 1294;
Best Local Similarity 66.7%: Pred. No. 1,22e+01:
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 126 LPQDVLSP 134
||: ||||
Oy 1 LPENNVLSP 9

RESULT 8

```

ENTRY S49775 #type complete  
 TITLE hypothetical protein YDR179c - yeast (Saccharomyces cerevisiae)  
 ALTERNATE\_NAMES hypothetical protein YD9395.12  
 ORGANISM #formal\_name Saccharomyces cerevisiae  
 DATE 13-Jan-1995 #sequence\_revision 10-Feb-1995 #text\_change 12-Dec-1997

ACCESSIONS S49775  
 REFERENCE S49764  
 #authors Murphy, L.; Harris, D.E.  
 #submission submitted to the EMBL Data Library, November 1994  
 #accession S49775  
 #molecule\_type DNA  
 #residues 1-162 #label MUR  
 #cross-references EMBL:Z46727; NID:g1289283; PID:e223643; PID:g1289294; MIPS:YDR179c

GENETICS  
 #map\_position 4R  
 SUMMARY #length 162 #molecular-weight 19476 #checksum 688

Query Match 71.6%; Score 48; DB 2; Length 162;  
 Best Local Similarity 60.0%; Pred. No. 1.89e+01;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 46 LPENITSL 55  
 1 LPENNVSPL 10

RESULT 9  
 ENTRY S52095 #type complete  
 TITLE tau-protein kinase (EC 2.7.1.135) homolog - common tobacco  
 ALTERNATE\_NAMES glycogen synthase kinase 3 homolog; protein kinase GSK-3  
 ORGANISM #molecule\_type protein kinase shaggy homolog  
 #formal\_name Nicotiana tabacum #common\_name common tobacco  
 DATE 14-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 08-Sep-1997

ACCESSIONS S52095; S42085  
 REFERENCE S52095  
 #authors Einzenberger, E.; Eller, N.; Heberle-Bors, E.; Vicente, O.  
 #journal Blochm. Biophys. Acta (1995) 1260:315-319  
 #title Isolation and expression during pollen development of a tobacco cDNA clone encoding a protein kinase homologous to shaggy/glycogen synthase kinase-3.

#accession S52095  
 #status preliminary  
 #molecule\_type mRNA  
 #residues 1-409 #label EIN  
 #cross-references EMBL:X77763; NID:g456355; PID:g456356

CLASSIFICATION #superfamily kinase-related transforming protein; protein kinase homolog  
 #keywords ATP; phosphotransferase; serine/threonine-specific protein kinase

FEATURES  
 71-332 #domain protein kinase homolog #label KIN  
 79-87 #region protein kinase ATP-binding motif  
 102 #active-site Lys #status predicted  
 SUMMARY #length 409 #molecular-weight 46308 #checksum 7036

Query Match 71.6%; Score 48; DB 2; Length 409;  
 Best Local Similarity 60.0%; Pred. No. 1.89e+01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 367 LPNGRVLPPL 376  
 1 LPENNVSPL 10

RESULT 10  
 ENTRY F69378 #type complete  
 TITLE conserved hypothetical protein AF1030 - Archaeoglobus fulgidus  
 ORGANISM #formal\_name Archaeoglobus fulgidus

DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Dec-1997

ACCESSIONS F69378  
 REFERENCE A69250  
 #authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gilm, M.; Hickey, E.R.; Peterson, J.D.; Richardson, D.L.; Kervatage, A.R.; Graham, D.E.; Kyriades, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Arlrich, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Nature (1997) 390:364-370  
 #title The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.  
 #cross-references GB:AE00782; TIGR:AF1030  
 #accession F69378  
 #status preliminary; nucleic acid sequence not shown; translation not shown

#molecule\_type DNA  
 #residues 1-527 #label KLE  
 #cross-references GB:AE00782; TIGR:AF1030  
 SUMMARY #length 527 #molecular-weight 58876 #checksum 1184

Query Match 71.6%; Score 48; DB 2; Length 527;  
 Best Local Similarity 50.0%; Pred. No. 1.89e+01;  
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 113 LADDVLSL 122  
 1 LPENNVSPL 10

RESULT 11  
 ENTRY S75175 #type complete  
 TITLE hypothetical protein sl1945 - Synechocystis sp. (PCC 6803)  
 ORGANISM #formal\_name Synechocystis sp.  
 #variety PCC 6803  
 DATE 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Sep-1997

ACCESSIONS S75175  
 REFERENCE S74322  
 #authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Nario, K.; Okumura, S.; Shampo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.

#journal DNA Res. (1996) 3:109-136  
 #title Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.

#cross-references MIM:97061201  
 #accession S75175  
 #status preliminary  
 #molecule\_type DNA  
 #residues 1-640 #label KAN  
 #cross-references EMBL:D90903; NID:g1652127; PID:d1017822; PID:g1652165  
 #note the nucleotide sequence was submitted to the EMBL Data Library, June 1996

SUMMARY #length 640 #molecular-weight 69327 #checksum 6213

Query Match 71.6%; Score 48; DB 2; Length 640;  
 Best Local Similarity 70.0%; Pred. No. 1.89e+01;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 580 LMDNVLPPL 589

OY 1 LPENNVLSPL 10

RESULT 12

ENTRY 545299 #type complete

TITLE hypothetical protein - Trypanosoma brucei

ORGANISM #formal\_name Trypanosoma brucei

DATE 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 06-Jan-1995

ACCESSIONS S45299

REFERENCE S45299

#authors Glauser, A.; Braun, R.

#journal Biochim. Biophys. Acta (1994) 1218:99-101

#title TUBIS, a fossilized retroposon in the tubulin gene cluster of Trypanosoma brucei.

#accession S45299

#status preliminary

#molecule\_type DNA

#residues 1-101 #label GLA

SUMMARY #cross-references EMBL:X62164

#length 101 #molecular-weight 11327 #checksum 995

Query Match 70.1%; Score 47; DB 2; Length 101;

Best Local Similarity 50.0%; Pred. No. 2.91e+01;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 9 VPQDTVLGPL 18

OY 1 LPENNVLSPL 10

RESULT 13

ENTRY JC4042 #type complete

TITLE lexa protein - Aeromonas hydrophila

ORGANISM #formal\_name Aeromonas hydrophila

DATE 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 08-Sep-1997

ACCESSIONS JC4042

REFERENCE JC4042

#authors Rierley, J.; Barde, J.

#journal Gene (1995) 154:71-75

#title Cloning, sequence and regulation of expression of the lexa gene of Aeromonas hydrophila.

#accession JC4042

#status preliminary

#molecule\_type DNA

#residues 1-207 #label RIE

SUMMARY #cross-references EMBL:X77263; NID:g840713; PID:g840714

GENETICS lexa

CLASSIFICATION #superfamily lexa repressor

KEYWORDS SOS response; transcription regulation

SUMMARY #length 207 #molecular-weight 22898 #checksum 4250

Query Match 70.1%; Score 47; DB 2; Length 207;

Best Local Similarity 70.0%; Pred. No. 2.91e+01;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 173 LPENNELSPI 182

OY 1 LPENNVLSPL 10

RESULT 14

ENTRY B69272 #type complete

TITLE conserved hypothetical protein AF0178 - Archaeoglobus fulgidus

ORGANISM #formal\_name Archaeoglobus fulgidus

DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Dec-1997

ACCESSIONS B69272

REFERENCE A69250

#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,

K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kervatage, A.R.; Graham, D.E.; Kyriades, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Nature (1997) 390:364-370

#title The complete genome sequence of the hyperthermophilic sulfate-reducing archaeon Archaeoglobus fulgidus.

#cross-references MUID:98049343

#accession B69272

#status preliminary; nucleic acid sequence not shown; translation not shown

##molecule\_type DNA

##residues 1-320 #label KLE

SUMMARY #cross-references GB:AE000782; TIGR:AF0178

#length 320 #molecular-weight 37301 #checksum 6049

Query Match 70.1%; Score 47; DB 2; Length 320;

Best Local Similarity 60.0%; Pred. No. 2.91e+01;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 96 LPXYTVITPL 105

OY 1 LPENNVLSPL 10

RESULT 15

ENTRY D69370 #type complete

TITLE acyl-CoA dehydrogenase (acd-6) homolog - Archaeoglobus fulgidus

ORGANISM #formal\_name Archaeoglobus fulgidus

DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Dec-1997

ACCESSIONS D69370

REFERENCE A69250

#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kervatage, A.R.; Graham, D.E.; Kyriades, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Nature (1997) 390:364-370

#title The complete genome sequence of the hyperthermophilic sulfate-reducing archaeon Archaeoglobus fulgidus.

#cross-references MUID:98049343

#accession D69370

#status preliminary; nucleic acid sequence not shown; translation not shown

##molecule\_type DNA

##residues 1-394 #label KLE

SUMMARY #cross-references GB:AE000782; TIGR:AF0964

#length 394 #molecular-weight 43800 #checksum 9015

Query Match 70.1%; Score 47; DB 2; Length 394;

Best Local Similarity 60.0%; Pred. No. 2.91e+01;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 224 LPENALIPPL 233

SUN SEP 13 10:55:23 1998

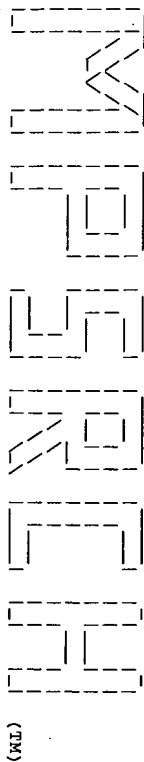
US-08-452-843-13.rpr

OY 1 LPENNVLSPL 10

Search completed: Fri Sep 11 13:25:41 1998  
Job time : 15 secs

**THIS PAGE BLANK (USPTO)**





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Sep 11 13:25:59 1998; Maspar time 2.28 Seconds  
Tubular output not generated. 109.870 Million cell updates/sec

Title: >US-08-452-843-13  
Description: (1-10) from US08452843.pep  
Perfect Score: 67  
Sequence: 1 LPENNIVSPL 10

Scoring table: PAM 150  
Gap 15

Searched: 69111 segs, 25083644 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swiss1

Statistics: Mean 23.945; Variance 25.262; scale 0.948

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID         | Description              | Pred. No. |
|------------|-------|-------------|--------|----|------------|--------------------------|-----------|
| 1          | 67    | 100.0       | 393    | 1  | P53_CERAE  | CELLULAR TUMOR ANTIGEN   | 8.11e-05  |
| 2          | 67    | 100.0       | 393    | 1  | P53_HUMAN  | CELLULAR TUMOR ANTIGEN   | 8.11e-05  |
| 3          | 60    | 89.6        | 314    | 1  | P53_SPEBE  | CELLULAR TUMOR ANTIGEN   | 5.97e-03  |
| 4          | 52    | 77.6        | 276    | 1  | P53_CANFA  | CELLULAR TUMOR ANTIGEN   | 5.76e-01  |
| 5          | 52    | 77.6        | 386    | 1  | P53_FELCA  | CELLULAR TUMOR ANTIGEN   | 9.88e-01  |
| 6          | 51    | 76.1        | 393    | 1  | P53_CRIGR  | CELLULAR TUMOR ANTIGEN   | 9.88e-01  |
| 7          | 51    | 76.1        | 396    | 1  | P53_MESAU  | CELLULAR TUMOR ANTIGEN   | 9.88e-01  |
| 8          | 51    | 76.1        | 1752   | 1  | RPE1_SCHPO | DNA-DIRECTED RNA POLYM   | 9.88e-01  |
| 9          | 50    | 74.6        | 785    | 1  | DMGA_ECOLI | ANAEROBIC DIMETHYL SUL   | 1.68e+00  |
| 10         | 47    | 70.1        | 207    | 1  | LEXA_AERYH | LEXA REPRESSOR (EC 3.4   | 7.93e+00  |
| 11         | 47    | 70.1        | 577    | 1  | YIIP_ECOLI | HYPOTHETICAL 66.6 KD P   | 7.93e+00  |
| 12         | 47    | 70.1        | 1124   | 1  | POL_FITV2  | POL POLYPROTEIN (PROTE   | 7.93e+00  |
| 13         | 46    | 68.7        | 266    | 1  | E1A_ADEL2  | EARLY E1A 29.5 KD PROT   | 1.31e+00  |
| 14         | 46    | 68.7        | 382    | 1  | P53_SHEEP  | CELLULAR TUMOR ANTIGEN   | 1.31e+00  |
| 15         | 46    | 68.7        | 386    | 1  | P53_BOVIN  | CELLULAR TUMOR ANTIGEN   | 1.31e+00  |
| 16         | 46    | 68.7        | 537    | 1  | IL2B_RAT   | INTERLEUKIN-2 RECEPTOR   | 1.31e+00  |
| 17         | 45    | 67.2        | 348    | 1  | YLM1_CAEEL | HYPOTHETICAL 41.0 KD P   | 2.13e+00  |
| 18         | 45    | 67.2        | 640    | 1  | PCOM_CHICK | PHOSPHONOENOLPYRUVATE CA | 2.13e+00  |
| 19         | 45    | 67.2        | 1124   | 1  | POL_FITV2  | POL POLYPROTEIN (PROTE   | 2.13e+00  |
| 20         | 45    | 67.2        | 1124   | 1  | POL_FITV2  | POL POLYPROTEIN (PROTE   | 2.13e+00  |
| 21         | 44    | 65.7        | 303    | 1  | PYR3_SCHRO | PROTEIN-TYROSINE PHOSP   | 3.45e+00  |
| 22         | 44    | 65.7        | 304    | 1  | CAH1_VARY  | CELL SURFACE-BINDING P   | 3.45e+00  |
| 23         | 44    | 65.7        | 304    | 1  | CAH1_VACCV | CELL SURFACE-BINDING P   | 3.45e+00  |

|    |    |      |      |   |             |                          |          |
|----|----|------|------|---|-------------|--------------------------|----------|
| 24 | 44 | 65.7 | 304  | 1 | CAH1_VACCV  | CELL SURFACE-BINDING P   | 3.45e+01 |
| 25 | 44 | 65.7 | 386  | 1 | RPA2_METVA  | DNA-DIRECTED RNA POLYM   | 3.45e+01 |
| 26 | 44 | 65.7 | 468  | 1 | KG3H_DICDI  | GLYCOCEN SYNTHASE KINA   | 3.45e+01 |
| 27 | 44 | 65.7 | 495  | 1 | ACCD_MYCTU  | PURITATIVE ACETYL-COENZY | 3.45e+01 |
| 28 | 44 | 65.7 | 502  | 1 | YMAO_MARPO  | HYPOTHEICAL 57.7 KD P    | 3.45e+01 |
| 29 | 44 | 65.7 | 783  | 1 | GCR_MOUSE   | GLUCOCORTICOID RECEPTO   | 3.45e+01 |
| 30 | 44 | 65.7 | 795  | 1 | GCR_RAT     | GLUCOCORTICOID RECEPTO   | 3.45e+01 |
| 31 | 44 | 65.7 | 806  | 1 | DMGA_HAEIN  | ANAEAROBIC DIMETHYL SUL  | 3.45e+01 |
| 32 | 43 | 64.2 | 173  | 1 | YBC1_ECOLI  | HYPOTHEICAL 19.5 KD P    | 5.52e+01 |
| 33 | 43 | 64.2 | 213  | 1 | R1I_METVA   | 50S RIBOSOMAL PROTEIN    | 5.52e+01 |
| 34 | 43 | 64.2 | 275  | 1 | RL2_BACST   | 50S RIBOSOMAL PROTEIN    | 5.52e+01 |
| 35 | 43 | 64.2 | 276  | 1 | RL2_BACST   | 50S RIBOSOMAL PROTEIN    | 5.52e+01 |
| 36 | 43 | 64.2 | 276  | 1 | RL2_THEMA   | 50S RIBOSOMAL PROTEIN    | 5.52e+01 |
| 37 | 43 | 64.2 | 353  | 1 | REON_HUMAN  | ZINC-FINGER PROTEIN NE   | 5.52e+01 |
| 38 | 43 | 64.2 | 391  | 1 | P53_RABIT   | CELLULAR TUMOR ANTIGEN   | 5.52e+01 |
| 39 | 43 | 64.2 | 483  | 1 | KG3A_RAT    | GLYCOCEN SYNTHASE KINA   | 5.52e+01 |
| 40 | 43 | 64.2 | 483  | 1 | KG3A_HUMAN  | GLYCOCEN SYNTHASE KINA   | 5.52e+01 |
| 41 | 43 | 64.2 | 521  | 1 | YMA1_XENLA  | IMPORTIN ALPHA-1 SUBUN   | 5.52e+01 |
| 42 | 43 | 64.2 | 712  | 1 | CNAC_HUMAN  | CAMP-DEPENDENT 3',5'-C   | 5.52e+01 |
| 43 | 43 | 64.2 | 774  | 1 | RRP3_INCUJ  | RNA-DIRECTED RNA POLYM   | 5.52e+01 |
| 44 | 43 | 64.2 | 774  | 1 | RRP3_INCBSE | RNA-DIRECTED RNA POLYM   | 5.52e+01 |
| 45 | 43 | 64.2 | 2171 | 1 | C10C_RABIT  | DIHYDROPYRIDINE-SENSIT   | 5.52e+01 |

# ALIGNMENTS

| RESULT | ID         | Score  | Match | Length | DB         | ID                      | Description | Pred. No. |
|--------|------------|--------|-------|--------|------------|-------------------------|-------------|-----------|
| 1      | P53_CERAE  | 100.08 | 393   | 1      | P53_CERAE  | CELLULAR TUMOR ANTIGEN  | 8.11e-05    |           |
| 2      | P53_HUMAN  | 100.08 | 393   | 1      | P53_HUMAN  | CELLULAR TUMOR ANTIGEN  | 8.11e-05    |           |
| 3      | P53_SPEBE  | 89.6   | 314   | 1      | P53_SPEBE  | CELLULAR TUMOR ANTIGEN  | 5.97e-03    |           |
| 4      | P53_CANFA  | 77.6   | 276   | 1      | P53_CANFA  | CELLULAR TUMOR ANTIGEN  | 5.76e-01    |           |
| 5      | P53_FELCA  | 77.6   | 386   | 1      | P53_FELCA  | CELLULAR TUMOR ANTIGEN  | 9.88e-01    |           |
| 6      | P53_CRIGR  | 76.1   | 393   | 1      | P53_CRIGR  | CELLULAR TUMOR ANTIGEN  | 9.88e-01    |           |
| 7      | P53_MESAU  | 76.1   | 396   | 1      | P53_MESAU  | CELLULAR TUMOR ANTIGEN  | 9.88e-01    |           |
| 8      | RPE1_SCHPO | 74.6   | 785   | 1      | RPE1_SCHPO | DNA-DIRECTED RNA POLYM  | 9.88e-01    |           |
| 9      | DMGA_ECOLI | 70.1   | 207   | 1      | DMGA_ECOLI | ANAEAROBIC DIMETHYL SUL | 1.68e+00    |           |
| 10     | LEXA_AERYH | 70.1   | 577   | 1      | LEXA_AERYH | LEXA REPRESSOR (EC 3.4  | 7.93e+00    |           |
| 11     | YIIP_ECOLI | 70.1   | 1124  | 1      | YIIP_ECOLI | HYPOTHEICAL 66.6 KD P   | 7.93e+00    |           |
| 12     | POL_FITV2  | 68.7   | 266   | 1      | POL_FITV2  | POL POLYPROTEIN (PROTE  | 7.93e+00    |           |
| 13     | EIA_ADEL2  | 68.7   | 382   | 1      | EIA_ADEL2  | EARLY EIA 29.5 KD PROT  | 1.31e+01    |           |
| 14     | P53_SHEEP  | 68.7   | 386   | 1      | P53_SHEEP  | CELLULAR TUMOR ANTIGEN  | 1.31e+01    |           |
| 15     | P53_BOVIN  | 68.7   | 386   | 1      | P53_BOVIN  | CELLULAR TUMOR ANTIGEN  | 1.31e+01    |           |
| 16     | IL2B_RAT   | 67.2   | 537   | 1      | IL2B_RAT   | INTERLEUKIN-2 RECEPTOR  | 1.31e+01    |           |
| 17     | YLM1_CAEEL | 67.2   | 348   | 1      | YLM1_CAEEL | HYPOTHEICAL 41.0 KD P   | 2.13e+01    |           |
| 18     | PCOM_CHICK | 67.2   | 640   | 1      | PCOM_CHICK | PHOSPHONOPIRYVATE CA    | 2.13e+01    |           |
| 19     | POL_FITV2  | 67.2   | 1124  | 1      | POL_FITV2  | POL POLYPROTEIN (PROTE  | 2.13e+01    |           |
| 20     | POL_FITV2  | 67.2   | 1124  | 1      | POL_FITV2  | POL POLYPROTEIN (PROTE  | 2.13e+01    |           |
| 21     | PRP3_SCHPO | 65.7   | 303   | 1      | PRP3_SCHPO | PROTEIN-TYROSINE PHOSP  | 3.45e+01    |           |
| 22     | CAH1_VARY  | 65.7   | 304   | 1      | CAH1_VARY  | CELL SURFACE-BINDING P  | 3.45e+01    |           |
| 23     | CAH1_VACCV | 65.7   | 304   | 1      | CAH1_VACCV | CELL SURFACE-BINDING P  | 3.45e+01    |           |

Query Match 100.0%; Score 67; DB 1; Length 393;

Best Local Similarity 100.0%; Pred. No. 8.11e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSPL 35  
1 LPENNVLSPL 10

RESULT 2  
ID P53 HUMAN STANDARD; PRT: 393 AA.  
AC P04637;  
DT 13-AUG-1987 (REL: 05, CREATED)  
DT 01-MAR-1989 (REL: 10, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL: 35, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
GN TP53.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85230577.  
RA ZAKUT-HOORI R., BIENZ-TADMOR B., GIVOL D., OREN M.;  
ENBO J. 4:1251-1255(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87064416.  
RA LAMB P., CRAWFORD L.;  
MOL. CELL. BIOL. 6:1379-1385(1986).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85267676.  
RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;  
MOL. CELL. BIOL. 5:1601-1610(1985).  
RN [4]  
RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
RX MEDLINE: 87089826.  
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
ROTTER V.;  
MOL. CELL. BIOL. 6:4650-4656(1986).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 89108008.  
RA BUCHANAN V.L., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P.,  
GEOBGIEV G.P.;  
GENE 70:245-252(1988).  
RN [6]  
RP SEQUENCE OF 101-393 FROM N.A.  
RX MEDLINE: 85126934.  
RA MATLASHENSKI G., LAMB P., PIW D., PEACOCK J., CRAWFORD L.,  
BENCHIMOL S.;  
EMBO J. 3:3257-3262(1984).  
RN [7]  
RP NUCLEAR LOCALIZATION SIGNAL.  
RX MEDLINE: 90191730.  
RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
ONCOGENE 5:423-426(1990).  
RN [8]  
RP STRUCTURE BY NMR OF 319-360.  
RX MEDLINE: 94294808.  
RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
APPELLA E., GRONENBORN A.M.;  
SCIENCE 265:386-391(1994).  
RN [9]  
RP STRUCTURE BY NMR OF 325-355.  
RX MEDLINE: 95292092.  
RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
MOL. STRUCT. BIOL. 1:877-890(1994).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
RX MEDLINE: 94294806.  
RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
SCIENCE 265:346-355(1994).

RN [11]  
RP REVIEW.  
RX MEDLINE: 94090335.  
RA HARRIS C.C.;  
SCIENCE 262:1980-1981(1993).  
RN [12]  
RP REVIEW ON VARIANTS.  
RX MEDLINE: 91289156.  
RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
SCIENCE 253:49-53(1991).  
RN [13]  
RP REVIEW ON VARIANTS.  
RX MEDLINE: 96271983.  
RA DE VRIES E.M.G., RICCIE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
LIAO D., SOUSSI T., KOVACH J.S., SOMMER S.S.;  
HUM. MUTAT. 7:202-213(1996).  
RN [14]  
RP VARIANT ARG-72.  
RX MEDLINE: 91153807.  
RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
HUM. GENET. 86:369-370(1991).  
RN [15]  
RP VARIANT LFS THR-133.  
RX MEDLINE: 92034774.  
RA LAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
CANCER RES. 51:6385-6387(1991).  
RN [16]  
RP VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
RX MEDLINE: 91057657.  
RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
KIM D.H., KASSEL J., GRAY M.A., BISCHOF F.Z., TAINSKY M.A.,  
FRIEND S.H.;  
SCIENCE 250:1233-1238(1990).  
RN [17]  
RP VARIANT LFS ASP-245.  
RX MEDLINE: 91080929.  
RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
NATURE 348:747-749(1990).  
RN [18]  
RP VARIANT LFS LEU-272.  
RX MEDLINE: 92147883.  
RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
KNUTSEN T., MINNA J.D.;  
J. CLIN. INVEST. 89:640-647(1992).  
RN [19]  
RP VARIANTS LFS HIS-273 AND VAL-325.  
RX MEDLINE: 92228023.  
RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
GEBHARDT M.C., ANDERSEN T.I., BORRESSEN A.-L., LI F.P., GARBER J.,  
STRONG L.C.;  
NEW ENGL. J. MED. 326:1309-1315(1992).  
RN [20]  
RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
RX MEDLINE: 90295284.  
RA BARTER J., IGGO R., GANNON J., LANE D.P.;  
ONCOGENE 5:893-899(1990).  
RN [21]  
RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
RX MEDLINE: 91017544.  
RA RODRIGUES N.R., ROMAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
GANNON J.V., LANE D.P.;  
PROC. NATL. ACAD. SCI. U.S.A. 87:7555-7559(1990).  
RN [22]  
RP VARIANTS COLORECTAL CANCER MUTATIONS.  
RX MEDLINE: 91282784.  
RA ISHIOKA C., SATO T., GAMOH M., SUZUKI T., SHIBATA H., KANAMARU R.,  
WAKUI A., YAMAZAKI T.;  
BIOCHEM. BIOPHYS. RES. COMMUN. 177:901-906(1991).  
RN [23]  
RP VARIANTS OESOPHAGUS TUMORS L-152; A-155; H-175; F-176 AND H-273.  
RX MEDLINE: 91330175.  
RA CASSON A.G., MUKHOPADHYAY T., CLEARY K.R., RO J.Y., LEVIN B.,

RA ROTH J.A.;  
 RL CANCER RES. 51:4495-4499(1991).  
 RN [24]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN CHINA.  
 RX MEDLINE; 91187113.  
 RA HSU I.C., METCALF R.A., SUN T., WELSH J.A., WANG N.J., HARRIS C.C.;  
 RL NATURE 350:427-428(1991).  
 RN [25]  
 RP VARIANTS HEPATOCELLULAR CARCINOMAS MUTATIONS IN SOUTH AFRICA.  
 RX MEDLINE; 91187114.  
 RA BRESSAC B., KEM M., WANDS J., OZTURK M.;  
 RL NATURE 350:429-431(1991).  
 RN [26]  
 RP VARIANTS IN ANGIOGENITAL CARCINOMAS.  
 RX MEDLINE; 93010989.  
 RA CROOK T., VOUSDEN K.H.;  
 RL EMO J. 11:3935-3940(1992).  
 RN [27]  
 RP VARIANT WITH DUPLICATION OF PRO-177-HIS-178.  
 RX MEDLINE; 93265016.  
 RA BHATTA K., GUTERREZ M.I., MAGRATH I.T.;  
 RL HUM. MOL. GENET. 1:207-208(1992).  
 RN [28]  
 RP VARIANTS IN BURKITT'S LYMPHOMAS.  
 RX MEDLINE; 93064692.  
 RA DOTHU A., DEBUIRE B., ROMANO J.W., EHRHART J.C., FISCELLA M., MAY E.,  
 RL APPELLA E., MAY P.;  
 ONCOGENE 7:2161-2167(1992).  
 RN [29]  
 RP VARIANT NASOPHARYNGEAL CARCINOMA THR-280.  
 RX MEDLINE; 92335329.  
 RA SUN Y., HEGAMER G., HENG Y.-J., HILDESHEIM A., CHEN J.-Y., CAO Y.,  
 RA YAO K.-T., COLBRIN N.H.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:6516-6520(1992).  
 RN [30]  
 RP VARIANTS IN COLON TUMORS.  
 RX MEDLINE; 93330562.  
 RA HAMELIN R., JEGO N., LAURENT-PUIG P., VIDAUD M., THOMAS G.;  
 RL ONCOGENE 8:2213-2220(1993).  
 RN [31]  
 RP CHARACTERIZATION OF VARIANT ALA-143.  
 RX MEDLINE; 94283378.  
 RA ZHANG W., GHO X.-Y., HU G.-Y., LIU W.-B., SHAY J.W., DEISSEROTH A.B.;  
 RL EMO J. 13:2535-2544(1994).  
 RN [32]  
 RP VARIANTS LFS HIS-175; ARG-193; GLN-248; CYS-273 AND TYR-275.  
 RX MEDLINE; 95193787.  
 RA FREBOURG T., BARBIER N., YAN Y.-X., GARBER J.E., DREYFUS M.,  
 RA FRAUDENI J.F., JR., LI F.P., FRIEND S.H.;  
 RL AM. J. HUM. GENET. 56:608-615(1995).  
 RN [33]  
 RP VARIANT LFS HIS-175.  
 RX MEDLINE; 96423319.  
 RA VARLEY J.M., MCGROWN G., THORNCROFT M., TRICKER K.J., TEARE M.D.,  
 RA SANTIBANEZ-MORE M.F., HOLSTON R.S., MARTIN J., BIRCH J.M.,  
 RA EVANS D.G.R.;  
 RL J. MED. GENET. 32:942-945(1995).  
 RN [34]  
 RP VARIANTS BA PHE-176; SER-245; TRP-248; TRP-282 AND GLN-286.  
 RX MEDLINE; 96233927.  
 RA AUDREZET M.-P., ROBASKIEWICZ M., MERCIER B., NOUSBAUM J.-B.,  
 RA HARDY E., BAIL J.-P., VOLANT A., LOZACH P., GOUEROU H., FEREC C.;  
 RL HUM. MUTAT. 7:109-113(1996).

Note: remainder of annotations omitted.

Query Match 100.0%; Score 67; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8,11e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 26. LPENNVLSPL 35  
 1 LPENNVLSPL 10

RESULT 3  
 ID P53\_SPEBE STANDARD; PRT; 314 AA.  
 AC 064662;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN TP53.  
 OS SPERMOPHILUS BEECHERI (BEECHER GROUND SQUIRREL).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=THYMUS;  
 RX MEDLINE; 95007566.  
 RA RIVKINA M.B., CULLEN J.M., ROBINSON W.S., MARION P.L.;  
 RL CANCER RES. 54:5430-5437(1994).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC DR EMBL; U43902; G1165312; -  
 CC DR PROSITE; P500348; P53; 1.  
 CC KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 CC NUCLEAR PROTEIN; PHOSPHORYLATION.  
 CC FT NON\_TER 1  
 CC FT DOMAIN 289 301 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT NON\_TER 314 314  
 CC FT NON\_TER 314 314  
 CC SQ SEQUENCE 314 AA; 34618 MW; D07F433B CRC32;  
 Db 6 LPENNVLSPL 15  
 1 LPENNVLSPL 10

Query Match 89.6%; Score 60; DB 1; Length 314;  
 Best Local Similarity 90.0%; Pred. No. 5,97e-03;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 6 LPENNVLSPL 15  
 1 LPENNVLSPL 10

RESULT 4  
 ID P53\_CANFA STANDARD; PRT; 276 AA.  
 AC 029357;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN TP53.  
 OS CANIS FAMILIARIS (DOG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; CARNIVORA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BEAGLE;  
 RX MEDLINE; 95323915.  
 RA KRAEDEL S.A., PAZZI K.A., MADEWELL B.R.;  
 RL CANCER LETT. 92:181-186(1995).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A

CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC EMBL: S77819; G1000577; -  
 CC PROSITE: PS00348; P53; 1.  
 CC ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 CC NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 CC FT NON-TER 1 35 ASP/GLU-RICH (ACIDIC).  
 CC FT DOMAIN 1 276  
 CC FT NON-TER 276  
 CC SEQUENCE 276 AA; 30466 MW; 8C97AE44 CRC32;  
 SQ  
 Query Match 77.6%; Score 52; DB 1; Length 276;  
 Best Local Similarity 88.9%; Pred. No. 5.76e-01;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 2 LPPNNVLS 10  
 QY 1 LPPNNVLS 9  
 RESULT 5  
 ID P53\_FELCA STANDARD; PRT: 386 AA.  
 AC P41685;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS FELIS SILVESTRIS CATUS (CAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; CARNIVORA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LYMPH NODE;  
 RX MEDLINE: 94333960.  
 RA OKUDA M., UMEBA A., SAKAI T., OHASHI T., MOMOI Y., YOUN H.Y.,  
 RA WATARI T., GOTTSUKA R., TSUJIMOTO H., HASEGAWA A.;  
 RL INT. J. CANCER 58:602-607(1994).  
 RN [2]  
 RP SEQUENCE OF 34-354 FROM N.A.  
 RX MEDLINE: 94114699.  
 RA OKUDA M., UMEBA A., MATSUMOTO Y., MOMOI Y., WATARI T., GOTTSUKA R.,  
 RA O'BRIEN S.J., TSUJIMOTO H., HASEGAWA A.;  
 RL J. VET. MED. SCI. 55:801-805(1993).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC EMBL: D26608; G538225; -  
 CC EMBL: D16460; G575528; -  
 CC PROSITE: PS00348; P53; 1.  
 CC ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 CC NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 CC FT NON-TER 1 35 ASP/GLU-RICH (ACIDIC).  
 CC FT DOMAIN 1 276  
 CC FT NON-TER 276  
 CC SEQUENCE 276 AA; 30466 MW; 8C97AE44 CRC32;  
 SQ

FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 285 285 K -> R (IN REF. 2).  
 SQ SEQUENCE 386 AA; 42692 MW; DEC7132A CRC32;  
 Query Match 77.6%; Score 52; DB 1; Length 386;  
 Best Local Similarity 88.9%; Pred. No. 5.76e-01;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 26 LPPNNVLS 34  
 QY 1 LPPNNVLS 9  
 RESULT 6  
 ID P53\_CRIGR STANDARD; PRT: 393 AA.  
 AC 009185; G64397; P97258; P97788;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR P53.  
 OS CRICETULUS GRISEUS (CHINESE HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CHAUNG W., MI L.J., BOORSTEIN R.J.;  
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 97183659.  
 RA LEE H., LARNER J.M., HAMLIN J.L.;  
 RL GENE 184:177-183(1997).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC EMBL: Y08900; E303876; -  
 CC EMBL: Y08901; E303863; -  
 CC DR EMBL: U50395; G1842230; -  
 CC PROSITE: PS00348; P53; 1.  
 CC ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 CC NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 CC FT NON-TER 1 74 ASP/GLU-RICH (ACIDIC).  
 CC FT DOMAIN 1 316  
 CC FT DOMAIN 316 390  
 CC FT DOMAIN 311 323  
 CC FT MOD\_RES 392 392 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT VARIANT 133 133 L -> Q (IN CELL LINE V79-4).  
 CC FT VARIANT 135 135 C -> W (IN CELL LINE V79-4).  
 CC FT CONFLICT 103 103 Y -> F (IN REF. 2).  
 SQ SEQUENCE 393 AA; 43378 MW; 402EB149 CRC32;  
 Query Match 76.1%; Score 51; DB 1; Length 393;  
 Best Local Similarity 80.0%; Pred. No. 9.88e-01;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Db 26 LPPNNVLS 35

OY 1 LPENNVLSP 10  
 RESULT 7  
 ID P53\_MESAU STANDARD: PRT: 396 AA.  
 AC 000366; P97276;  
 DT 01-DEC-1992 (REL. 24, CREATED)  
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN 11P53.  
 OS MESOCRICETUS AURATUS (GOLDEN HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SYRIAN; TISSUE-KIDNEY;  
 RX MEDLINE; 92210007.  
 RA LEGROS Y., MCINTYRE P., SOUSSI T.;  
 RL GENE 112:247-250(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA HOU E.W., WISEMAN R.;  
 RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND P53 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 CC EMBL; M75144; G191415;  
 DR EMBL; U07182; G473579;  
 DR PIR; JH0633; JH0633.  
 DR HSSP; P04637; 1PES.  
 DR PROSITE; P800348; P53; 1.  
 KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 77 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 78 153 HYDROPHOBIC.  
 FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
 FT INTERACTION WITH DNA.  
 FT DOMAIN 314 326 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 395 395 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 188 188 G -> S (IN REF. 2).  
 SQ SEQUENCE 396 AA; 43631 MW; C2688ADE CRC32;  
 Query Match 76.1%; Score 51; DB 1; Length 396;  
 Best Local Similarity 80.0%; Pred. No. 9,88e-01;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Db 26 LPENNVLSP 35  
 OY 1 LPENNVLSP 10  
 RESULT 8  
 ID RPBL\_SCHPO STANDARD: PRT: 1752 AA.  
 AC P36594;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6)  
 DE (RNA POLYMERASE II SUBUNIT 1).

GN RPBL.  
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972;  
 RX MEDLINE; 9187661.  
 RA AZUMA Y., YAMAGISHI M., UESHIMA R., ISHITAMA A.;  
 RL NUCLEIC ACIDS RES. 19:461-468(1991).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 10 DIFFERENT SUBUNITS.  
 CC THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE FOUND IN  
 CC EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA PRECURSOR,  
 CC POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE III FOR 5S  
 CC AND TRNA GENES.  
 CC -1- SIMILARITY: TO OTHER RNAP II LARGE SUBUNITS.  
 DR EMBL; X56564; G5055;  
 DR PIR; S26849; S26849.  
 DR PROSITE; P800115; RNA\_POL\_II\_REPEAT; 24.  
 KW DNA-DIRECTED RNA POLYMERASE; TRANSCRIPTION; ZINC; REPEAT;  
 FT ZN\_FING 69 85 C2H2-TYPE (POTENTIAL).  
 FT DOMAIN 1554 1752 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 SQ SEQUENCE 1752 AA; 194161 MW; B7CFE872 CRC32;  
 Query Match 76.1%; Score 51; DB 1; Length 1752;  
 Best Local Similarity 70.0%; Pred. No. 9,88e-01;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 202 LPENRLSP 211  
 OY 1 LPENNVLSP 10  
 RESULT 9  
 ID DMSA\_ECOLI STANDARD: PRT: 785 AA.  
 AC P18775;  
 DT 01-NOV-1990 (REL. 16, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ANAEROBIC DIETHYL SULFOXIDE REDUCTASE CHAIN A PRECURSOR (EC 1.-.-.)  
 DE (DMSO REDUCTASE).  
 GN DMSA.  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 CC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 17-22.  
 RC STRAIN-K12 / C600;  
 RX MEDLINE; 89096500.  
 RA BILIOUS P.T., COLE S.T., ANDERSON W.F., WEINER J.H.;  
 RL MOL. MICROBIOL. 2:785-795(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE; 97061202.  
 RA OSHIMA T., AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A.,  
 RA IKEMOTO K., INADA T., ITOH T., KAJIHARA M., KANAI K., KASHIMOTO K.,  
 RA KIMURA S., KITAHARA M., MAKINO K., MASUDA S., MITI T., MIZOBUCHI K.,  
 RA MORI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHITO Y., SAITO N.,

RA SAMPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y.,  
 RA YANO M., HORIUCHI T.,  
 RL DNA RES. 3:137-155(1996).  
 RN (4)  
 RP. MUTAGENESIS.  
 RX MEDLINE: 9417175.  
 RA TRIEBER C.A., ROTHERY R.A., WEINER J.H.,  
 RL J. BIOL. CHEM. 269:7103-7109(1994).  
 CC -1- FUNCTION: TERMINAL REDUCTASE DURING ANAEROBIC GROWTH ON  
 VARIOUS SULFOXIDE AND N-OXIDE COMPOUNDS. ALLOWS E. COLI TO GROW  
 ANAEROBICALLY ON ME(2)SO AS RESPIRATORY OXIDANT.  
 CC -1- CATALYTIC ACTIVITY: REDUCES VARIOUS N-OXIDE AND SULFOXIDE  
 COMPOUNDS INCLUDING TRIMETHYLAMINE N-OXIDE.  
 CC -1- COFACTOR: MOLYBDENUM (MOLYBDOPTERIN); MAY BIND A 4FE-4S CLUSTER.  
 CC -1- SUBUNIT: HOMODIMER. THE COMPLEX CONSIST OF THREE SUBUNITS: DMSA,  
 THE REDUCTASE; DMSB, AN ELECTRON TRANSFER PROTEIN, AND DMSB,  
 A MEMBRANE ANCHOR PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF THE MEMBRANE.  
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC MOLYBDOPTERIN-CONTAINING  
 OXIDOREDUCTASES.  
 CC EMBL: J03412; G145755;  
 CC EMBL: AE000191; G1787121;  
 CC EMBL: D90727; G1651421;  
 CC PTR: S03785; S03785.  
 DR ECGENE: EG10232; DMSA.  
 DR PROSITE: PS00551; MOLYBDOPTERIN\_PROK\_1; 1.  
 DR PROSITE: PS00490; MOLYBDOPTERIN\_PROK\_2; 1.  
 DR PROSITE: PS00932; MOLYBDOPTERIN\_PROK\_3; 1.  
 KM OXIDOREDUCTASE; SIGNAL; MOLYBDENUM; 4FE-4S; IRON-SULFUR.  
 FT SIGNAL  
 FT CHAIN 1  
 FT 17 785 ANAEROBIC DIMETHYL SULFOXIDE REDUCTASE  
 FT METAL 34 34 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 38 38 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 42 42 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 75 75 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 SQ SEQUENCE 785 AA; 87449 MW; 9C1ADEB0 CRC32;  
 Query Match 74.6%; Score 50; DB 1; Length 785;  
 Best Local Similarity 60.0%; Pred. No. 1.68e+00;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Db 637 LPEGVIDPL 646  
 OY 1 LPENNVLSPL 10  
 RESULT 10  
 ID LEXA\_AERHY STANDARD; PRT; 207 AA.  
 AC Q44069;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE LEXA REPRESSOR (EC 3.4.21.88).  
 GN LEXA.  
 OS AEROMONAS HYDROPHILA.  
 OS PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC VIBRIONACEAE.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 7466;  
 RX MEDLINE: 95172406.  
 RA RIERA J., BARBE J.,  
 RL GENE 154:71-75(1995).  
 CC -1- FUNCTION: REPRESSSES A NUMBER OF GENES INVOLVED IN THE RESPONSE TO  
 DNA DAMAGE (SOS RESPONSE), INCLUDING RECA AND LEXA. BINDS TO A  
 16 BP PALINDROMIC SEQUENCE. IN THE PRESENCE OF SINGLE-STRANDED  
 DNA, RECA INTERACTS WITH LEXA CAUSING AN AUTOCATALYTIC CLEAVAGE  
 WHICH DISRUPTS THE DNA-BINDING PART OF LEXA, LEADING TO  
 DEREGULATION OF THE SOS REGULATION AND EVENTUALLY DNA REPAIR  
 (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 84-ALA-1-GLY-85 BOND IN  
 REPRESSOR LEXA.

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S24 ALSO KNOWN AS THE  
 CC UMOD/LEXA FAMILY.  
 DR EMBL: X77263; G840714;  
 KW TRANSCRIPTION REGULATION; REPRESSOR; DNA DAMAGE; HYDROLASE;  
 KM SERINE PROTEASE; DNA REPLICATION; SOS RESPONSE; DNA-BINDING.  
 FT DNA BIND 28 48 H-T-H MOTIF (BY SIMILARITY).  
 FT SITE 89 90 CLEAVAGE (AUTO-) (BY SIMILARITY).  
 FT ACT\_SITE 123 123 INVOLVED IN AUTO-CLEAVAGE  
 (BY SIMILARITY).  
 FT ACT\_SITE 161 161 INVOLVED IN AUTO-CLEAVAGE  
 (BY SIMILARITY).  
 FT SEQUENCE 207 AA; 22898 MW; A7AA3AB3 CRC32;  
 Query Match 70.1%; Score 47; DB 1; Length 207;  
 Best Local Similarity 70.0%; Pred. No. 7.93e+00;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 173 LPENNVLSPL 182  
 OY 1 LPENNVLSPL 10  
 RESULT 11  
 ID YJLP\_ECOLI STANDARD; PRT; 577 AA.  
 AC P32678;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 66.6 KD PROTEIN IN FRWD-PKC INTERGENIC REGION.  
 GN YJLP.  
 OS ESCHERICHIA COLI.  
 OS PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE: 94089392.  
 RA BLATTNER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J.,  
 RA DANIELS D.L.,  
 RL NUCLEIC ACIDS RES. 21:5408-5417(1993).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE YHBX/YHJW/YJDB FAMILY.  
 CC EMBL: U00006; G396302;  
 DR EMBL: AE000469; G1790392;  
 DR ECGENE: EG11914; YJLP.  
 DR HYPOTHETICAL PROTEIN; TRANSMEMBRANE; INNER MEMBRANE.  
 FT TRANSMEM 17 37 POTENTIAL.  
 FT TRANSMEM 44 64 POTENTIAL.  
 FT TRANSMEM 69 89 POTENTIAL.  
 FT TRANSMEM 119 139 POTENTIAL.  
 FT TRANSMEM 154 174 POTENTIAL.  
 SQ SEQUENCE 577 AA; 66609 MW; 293C63FA CRC32;  
 Query Match 70.1%; Score 47; DB 1; Length 577;  
 Best Local Similarity 70.0%; Pred. No. 7.93e+00;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 216 LPENNVLSPL 225  
 OY 1 LPENNVLSPL 10  
 RESULT 12  
 ID POL\_FIVT2 STANDARD; PRT; 1124 AA.  
 AC P31822;  
 DT 01-JUL-1993 (REL. 26, CREATED)  
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1993 (REL. 32, LAST ANNOTATION UPDATE)  
 DE POL POLYPROTEIN (PROTEASE (RETROPEPSIN) (EC 3.4.23.16); REVERSE  
 DE TRANSCRIPTASE (EC 2.7.7.49); DEOXYRIBIDINE 5'-TRIPHOSPHATE  
 DE NUCLEOTIDHYDROLASE (EC 3.6.1.23) (DUTPASE); RIBONUCLEASE H

DE (EC 3.1.26.4)).  
 GN POL.  
 OS FELINE IMMUNODEFICIENCY VIRUS (ISOLATE TM2) (FIV).  
 OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 OC LENTIVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91303718.  
 RA KIWOMASU T., MIYAZAWA T., FUKUSAWA M., MAKI N., HASEGAWA A., MIKAMI T.,  
 RA SAKURAGI J.I., FUKUSAWA M., MAKI N., HASEGAWA A., MIKAMI T.,  
 RA ADACHI A.;  
 RL J. VIROL. 65:4539-4542(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92198230.  
 RA MAKI N., MIYAZAWA T., FUKUSAWA M., HASEGAWA A., HAYAMI M., MIKI K.,  
 RA MIKAMI T.;  
 RL ARCH. VIROL. 123:29-45(1992).  
 CC -1- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE TO 5'-PHOSPHO-  
 CC MONESTER.  
 CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.  
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 DR EMBL: M59418; G323950.  
 DR PIR: B45557; B45557.  
 DR HSSP: P03369; IRLV.  
 DR PROSITE: P500141; ASP\_PROTEASE; 1.  
 DR POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;  
 KW RNA-DIRECTED DNA POLYMERASE; NUCLEOTIDE METABOLISM  
 FT CHAIN 34 156  
 FT CHAIN 157 2  
 FT CHAIN 7710 7846  
 FT CHAIN 847 1124  
 FT ACT\_SITE 68 68  
 FT ACT\_SITE 1124 126100  
 SO SEQUENCE 1124 AA; 126100 MW; 84930C98 CRC32;  
 Query Match 70.18; Score 47; DB 1; Length 1124;  
 Best Local Similarity 70.08; Pred. No. 7.93e+00;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 303 LPOGWLSP 312  
 QY 1 LPENNVLSP 10  
 RESULT 13  
 ID EIA\_ADEL2 STANDARD: PRT: 266 AA.  
 AC P03259;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE EARLY EIA 29.5 KD PROTEIN (CONTAINS: EARLY EIA 26 KD PROTEIN).  
 OS HUMAN ADENOVIRUS TYPE 12  
 OC VIRIDAE: DS-DNA NONENVELOPED VIRUSES; ADENOVIRIDAE; MASTADENOVIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 81022638.  
 RA SUGISAKI H., SUGIMOTO K., TAKANAMI M., SHIROKI K., SATO I.,  
 RA SHIMOJO H., SAMADA Y., UEMIZU Y., DESUGI S., FUJINAGA K.;  
 RL CELL 20:777-786(1980).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 81052432.  
 RA PERRICAUDET M., LE MOULLEC J.-M., TIOUJIS P., PETERSSON U.;  
 RL NATURE 288:174-176(1980).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94076430.  
 RA SPRENGEL J., SCHMITZ B., HEUS-NEITZEL D., ZOCK C., DOERFLER W.,  
 RL J. VIROL. 68:379-389(1994).  
 CC -1- FUNCTION: TRANS-ACTIVATES EARLY VIRAL PROMOTERS AND SOME CELLULAR

CC PROMOTERS.  
 DR EMBL: V000004; G58450;  
 DR EMBL: V000004; G58451;  
 DR EMBL: V000004; G58452;  
 DR EMBL: X73487; G313362;  
 DR PIR: A03828; A0A0G2.  
 DR PIR: S33928; S33928.  
 KW TRANSCRIPTION REGULATION; ACTIVATOR; EARLY PROTEIN;  
 KM ALTERNATIVE SPLICING; ZINC-FINGER; DNA-BINDING.  
 FT ZN\_FING 159 179  
 FT VARSPPLIC 31 244  
 FT VARSPPLIC 160 190  
 FT VARSPPLIC 191 234  
 FT VARSPPLIC 245 266  
 FT VARSPPLIC 259 266  
 FT VARSPPLIC 22 266  
 SO SEQUENCE 266 AA; 29691 MW; 09F4C3F9 CRC32;  
 Query Match 68.78; Score 46; DB 1; Length 266;  
 Best Local Similarity 70.08; Pred. No. 1.31e+01;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 79 LPEPVLSP 88  
 QY 1 LPENNVLSP 10  
 RESULT 14  
 ID P53 SHEEP STANDARD: PRT: 382 AA.  
 AC P51664;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS OVIS ARIES (SHEEP).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; ARTIODACTYLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BLOOD:  
 RX MEDLINE: 95352828.  
 RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
 RL DNA SEQ. 5:255-259(1995).  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND P53 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: STRONG, TO P53 IN OTHER HIGHER EUKARYOTES.  
 DR EMBL: X81705; G602357;  
 DR PROSITE: P500348; P53; 1.  
 KM ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KM NUCLEAR PROTEIN; PHOSPHORYLATION; APOPTOSIS.  
 FT DOMAIN 1 66  
 FT DOMAIN 300 312  
 FT MOD\_RES 381 381  
 FT MOD\_RES 42809 42809  
 SO SEQUENCE 382 AA; 42809 MW; 0CB99A00 CRC32;  
 Query Match 68.78; Score 46; DB 1; Length 382;  
 Best Local Similarity 77.88; Pred. No. 1.31e+01;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Db 26 LPENNVLSP 34  
 QY 1 LPENNVLSP 9

```

RESULT 15
ID P53_BOVIN STANDARD: PRT: 386 AA.
AC Q29628;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; ARTIODACTYLA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER.
RX MEDLINE: 95352829.
RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;
RL DNA SEQ. 5:261-264(1995).
RN [2]
RP SEQUENCE OF 13-386 FROM N.A.
RC STRAIN=HOLSTEIN; TISSUE=THYMUS;
RX MEDLINE: 96401400.
RA KOMORI H., ISHIGURO N., HORICHI M., SHINAGAWA M., AIDA Y.;
RL VET. IMMUNOL. IMMUNOPATHOL. 52:53-63(1996).
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
BAX AND BCL-2 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
IN MANY TYPES OF CANCER.
CC -1- SIMILARITY: STRONG. TO P53 IN OTHER HIGHER EUKARYOTES.
DR EMBL: X81704; G60233; -.
DR EMBL: D49825; G1729419; -.
DR PROSITE: PS00348; P53; 1.
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW NUCLEAR-PROTEIN; PHOSPHORYLATION; APOPTOSIS.
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD.RES 385 385 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 380 380 R -> T (IN REF. 2).
SQ SEQUENCE 386 AA: 43255 MW: 0322BF3D CRC32:

Query Match 68.7%; Score 46; DB 1; Length 386;
Best Local Similarity 77.8%; Pred. NO. 1.31e+01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

Db 26 LPENNLLSS 34  
 |||||:  
 OY 1 LPENNVLSP 9

Search completed: Fri Sep 11 13:26:06 1998  
 Job time : 7 secs.



(TM)

Distribution rights by Oxford Molecular Ltd

using Smith-Waterman algorithm

MasPar time 4.22 Seconds

values

```
:sp_phage 8:sp_plant
```

```
scale 0.951
```

1 score distribution.

| Description               | Pred. No |
|---------------------------|----------|
| P53 TRANSFORMATION SUP    | 6,40e-0  |
| CELLULAR TUMOR ANTIGEN    | 6,40e-0  |
| CELLULAR TUMOR ANTIGEN    | 6,40e-0  |
| CELLULAR TUMOR ANTIGEN    | 6,40e-0  |
| CELLULAR TUMOR ANTIGEN    | 6,40e-0  |
| P53 TRANSFORMATION SUP    | 6,40e-0  |
| P53 TRANSFORMATION SUP    | 6,40e-0  |
| CELLULAR TUMOR ANTIGEN    | 6,40e-0  |
| CELLULAR TUMOR ANTIGEN    | 6,40e-0  |
| P53 TRANSFORMATION SUP    | 6,40e-0  |
| P53 TRANSFORMATION SUP    | 6,40e-0  |
| CELLULAR TUMOR ANTIGEN    | 6,40e-0  |
| CELLULAR TUMOR ANTIGEN    | 6,40e-0  |
| P53 TRANSFORMATION SUP    | 6,40e-0  |
| P53 TRANSFORMATION SUP    | 6,40e-0  |
| CELLULAR TUMOR ANTIGEN    | 6,40e-0  |
| CELLULAR TUMOR ANTIGEN    | 6,40e-0  |
| MELANOMA KINA0272 (F)     | 9,91e-0  |
| HYPHENATED 31.1 KD P      | 9,91e-0  |
| HYPERMUTATIONAL 21.7 KD P | 1,72e+0  |
| SINAPTOGOLIN              | 2,35e+0  |
| SIMILAR TO THE HUMAN R    | 2,95e+0  |
| ANTIDRUG RESISTANCE R     | 2,95e+0  |

[illegible]

5 AA.  
 (CE UPDATE)  
 (TION UPDATE)  
 ).  
 AA; TETRAPODA, MAMMALIA,  
 USDEN K.H., CROOK T.;  
 CRC32;

```
DB 2; Length 245; .  
.40e-05;  
ches 0; Indels 0; Gaps 0
```

3 AA.  
CE UPDATE)  
TION UPDATE)  
A; TETRAPODA; MAMMALIA;  
SIDEN K. H., CROOK T.;

```

RL EMO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL: X60019; G506451; -.
DR PROSITE: PS00348; P53; 1.
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW NUCLEAR PROTEIN; PHOSPHORYLATION.
F1 VARIANT 213 213
F2 NON TER 393 393
SQ SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;

Query Match 100.0%; Score 67; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 6,40e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35
OY 1 LPENNVSPL 10

RESULT 3
ID 016811 PRELIMINARY; PRT; 393 AA.
AC 016811;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85126934.
RA MATLASHESKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,
RL EMO J. 3:3257-3262(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87064416.
RA LAMB P., CRAWFORD L.;
RL MOL. CELL. BIOL. 6:1379-1385(1986).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL: M13121; G386994; -.
DR EMBL: M13112; G386994; JOINED.
DR EMBL: M13113; G386994; JOINED.
DR EMBL: M13114; G386994; JOINED.
DR EMBL: M13115; G386994; JOINED.
DR EMBL: M13116; G386994; JOINED.
DR EMBL: M13117; G386994; JOINED.
DR EMBL: M13118; G386994; JOINED.
DR EMBL: M13119; G386994; JOINED.
DR EMBL: M13120; G386994; JOINED.
DR PROSITE: PS00348; P53; 1.
KW REPEAT; TUMOR ANTIGEN; ANTI-ONCOGENE; DNA-BINDING;
KW TRANSSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN;
KW PHOSPHORYLATION.
FT NON TER 393 393
SQ SEQUENCE 393 AA; 43698 MW; 3EA71431 CRC32;

Query Match 100.0%; Score 67; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 6,40e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 26 LPENNVSPL 35
OY 1 LPENNVSPL 10

RESULT 4
ID 016807 PRELIMINARY; PRT; 393 AA.
AC 016807;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RL EMO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL: X60011; G506435; -.
DR PROSITE: PS00348; P53; 1.
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW NUCLEAR PROTEIN; PHOSPHORYLATION.
FT VARIANT 193 193
FT NON TER 393 393
SQ SEQUENCE 393 AA; 43731 MW; 279BC9CB CRC32;

Query Match 100.0%; Score 67; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 6,40e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35
OY 1 LPENNVSPL 10

RESULT 5
ID 016808 PRELIMINARY; PRT; 393 AA.
AC 016808;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RL EMO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL: X60018; G506449; -.
DR PROSITE: PS00348; P53; 1.
KW ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW NUCLEAR PROTEIN; PHOSPHORYLATION.

```

FT VARIANT 163 163 H -> Y.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43627 MW: AFD8A9E3 CRC32:  
 Query Match  
 Best Local Similarity 100.0%; Score 67; DB 2; Length 393;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSPL 35  
 ||||||||  
 QY 1 LPENNVLSPL 10

RESULT 6  
 ID Q15087 PRELIMINARY; PRT: 393 AA.  
 AC Q15087:  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.:  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL: X60014; G506441; -.  
 FT VARIANT 237 237 I -> M.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43694 MW: 9BB81992 CRC32:  
 Query Match  
 Best Local Similarity 100.0%; Score 67; DB 2; Length 393;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7  
 ID Q15088 PRELIMINARY; PRT: 393 AA.  
 AC Q15088:  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST ANNOTATION UPDATE)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.:  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL: X60016; G506445; -.  
 FT VARIANT 238 238 Y -> C.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43713 MW: A01E1523 CRC32:  
 Query Match  
 Best Local Similarity 100.0%; Score 67; DB 2; Length 393;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8  
 ID Q16810 PRELIMINARY; PRT: 393 AA.  
 AC Q16810:  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.:  
 RL EMO J. 10:2879-2887(1991).  
 CC -I- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -I- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: X60020; G506453; -.  
 DR PROSITE: PS00348; P53; 1.  
 KW ANTI-ONCOGENE: DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION.  
 FT VARIANT 254 254 D -> N.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA: 43714 MW: 5F914579 CRC32:  
 Query Match  
 Best Local Similarity 100.0%; Score 67; DB 2; Length 393;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSPL 35  
 ||||||||  
 QY 1 LPENNVLSPL 10

RESULT 9  
 ID Q16848 PRELIMINARY; PRT: 393 AA.  
 AC Q16848:  
 DT 01-NOV-1996 (TREMBLERL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLERL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLERL. 05, LAST ANNOTATION UPDATE)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTER V.:  
 RL MOL. CELL. BIOL. 6:4650-4656(1986).  
 CC -I- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -I- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: M14694; G339814; -.  
 DR PROSITE: PS00348; P53; 1.  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ANTI-ONCOGENE; DNA-BINDING;  
 KW TRANSCRIPTION REGULATION; ACTIVATOR.  
 SQ SEQUENCE 393 AA: 43723 MW: DAYD302F CRC32:  
 Query Match  
 Best Local Similarity 100.0%; Score 67; DB 2; Length 393;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
|1111111111|  
QY 1 LPENNVSPL 10

RESULT 10  
ID 016535 PRELIMINARY; PRT: 393 AA.  
AC 016535;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
GN P53.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RL EMBL: X60017; G506447; -.  
DR EMBL: X60015; G506443; -.  
FT VARIANT 248 248 Q -> R.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA: 43684 MW: 239818A9 CRC32;

Query Match 100.0%; Score 67; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 6.40e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
|1111111111|  
QY 1 LPENNVSPL 10

RESULT 11  
ID 015086 PRELIMINARY; PRT: 393 AA.  
AC 015086;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
GN P53.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RL EMBL: X60013; G506439; -.  
FT VARIANT 246 246 T -> M.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA: 43682 MW: 943862A3 CRC32;

Query Match 100.0%; Score 67; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 6.40e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
|1111111111|  
QY 1 LPENNVSPL 10

RESULT 12  
ID 036006 PRELIMINARY; PRT: 391 AA.  
AC 036006;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE CELLULAR TUMOR ANTIGEN P53.

GN P53.  
OS MAMMATA MONAX.  
OG PLASMID PT7BLUE (R).  
OC UNCLASSIFIED.

RN [1]  
RP SEQUENCE FROM N.A.  
RA FEITELSON M.A., RANGANATHAN P.N., CLAYTON M.M., ZHANG S.M.;  
RL ONCOGENE 15:327-336(1997).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL: AJ001022; E351287; -.  
DR PROSITE: PS00348; P53; 1.  
KW PLASMID: ANTI-ONCOGENE; DNA-BINDING; TRANSCRIPTION REGULATION;  
KW ACTIVATOR; NUCLEAR PROTEIN; PHOSPHORYLATION  
SQ SEQUENCE 391 AA: 43468 MW: 95FA88F2 CRC32;

Query Match 89.6%; Score 60; DB 13; Length 391;  
Best Local Similarity 90.0%; Pred. No. 5.20e-03;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
|1111111111|  
QY 1 LPENNVSPL 10

RESULT 13  
ID 092560 PRELIMINARY; PRT: 726 AA.  
AC 092560;  
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE MYELOBLAST KINA0272 (FRAGMENT).  
GN KINA0272.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE-BRAIN;  
RA MEDLINE: 97191544.  
RA NAGASE T., SERI N., ISHIKAWA K., OHIRA M., KANARABAYASI Y.,  
RA OHARA O., TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;  
RL DNA RES. 3:321-329(1996).  
DR EMBL: D87462; D1014091; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 726 AA: 79988 MW: 0ED25A27 CRC32;

Query Match 79.1%; Score 53; DB 2; Length 726;  
Best Local Similarity 70.0%; Pred. No. 3.23e-01;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 561 LAEDGYLSP 570  
|1111111111|  
QY 1 LPENNVSPL 10

RESULT 14  
ID 004681 PRELIMINARY; PRT: 161 AA.  
AC 004681;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE PT15.  
OS LYCOPERSICON ESCULENTUM (TOMATO).  
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;  
OC SOLANALES; SOLANACEAE.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97357308.  
 RA ZHOU J., TANG X., MARTIN G.B.;  
 RL EMBL; 16:3207-3218(1997).  
 DR EMBL; U89256; G2213783; -.  
 SQ SEQUENCE 161 AA; 18051 MW; 9ADA5570 CRC32;

Query Match 76.1%; Score 51; DB 8; Length 161;  
 Best Local Similarity 70.0%; Pred. No. 9.91e-01;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 43 LPRNNILRPL 52  
 |||||  
 OY 1 LPENNVL SPL 10

RESULT 15  
 ID 006279 PRELIMINARY; PRT; 303 AA.  
 AC 006279;  
 DT 01-JUL-1997 (TREMBL.REL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBL.REL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1997 (TREMBL.REL. 04, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 31.1 KD PROTEIN.  
 GN MTCY07H7B.19.  
 OS MYCOBACTERIUM TUBERCULOSIS.  
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA DEVLIN K., CHURCHER C.M.;  
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA BARRELL B.G., RAJANBREM M.A.;  
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE; 96181548.  
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,  
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
 RA COLE S.T.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).  
 DR EMBL; 295557; E316966; -.  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 303 AA; 31104 MW; 4FAFF1EC CRC32;

Query Match 76.1%; Score 51; DB 9; Length 303;  
 Best Local Similarity 60.0%; Pred. No. 9.91e-01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 55 LPDTPVLPLPL 64  
 |||||  
 OY 1 LPENNVL SPL 10

Search completed: Fri Sep 11 13:26:49 1998  
 Job time : 26 secs.

**THIS PAGE BLANK (USPTO)**

# RELEASE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch.p protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:20:52 1998; Maspar time 2.67 Seconds  
54.591 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-12  
Description: (1-9) from US08452843.pep  
Perfect Score: 69  
Sequence: 1 MPRGVVTL 9

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-genseq32

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 16.464; Variance 48.375; scale 0.340

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description            | Pred. No. |
|------------|-------|-------------|--------|----|--------|------------------------|-----------|
| 1          | 69    | 100.0       | 9      | 18 | R89373 | B7 naturally processed | 3.38e-01  |
| 2          | 48    | 71.0        | 697    | 29 | W46517 | Saccharomyces cerevis  | 6.92e-01  |
| 3          | 48    | 69.6        | 1155   | 14 | R78167 | Mouse alpha-d subunit  | 8.91e-01  |
| 4          | 48    | 69.6        | 1155   | 26 | W23060 | Mouse beta 2 integrin  | 8.91e-01  |
| 5          | 48    | 69.6        | 1161   | 14 | R78168 | Mouse alpha-d subunit  | 8.91e-01  |
| 6          | 48    | 69.6        | 1161   | 26 | W23061 | Mouse beta 2 integrin  | 8.91e-01  |
| 7          | 47    | 68.1        | 97     | 25 | W32428 | Mycobacterium tubercu  | 1.14e-02  |
| 8          | 47    | 68.1        | 97     | 25 | W32360 | Mycobacterium tubercu  | 1.14e-02  |
| 9          | 47    | 68.1        | 151    | 24 | W13653 | Herpesvirus salmistr   | 1.14e-02  |
| 10         | 47    | 68.1        | 151    | 14 | R76571 | Herpesvirus ORF13 PRO  | 1.14e-02  |
| 11         | 47    | 68.1        | 151    | 19 | W02387 | HVS13 (viral homology  | 1.14e-02  |
| 12         | 47    | 68.1        | 229    | 12 | R70701 | Recombinant DNA-ase-B  | 1.14e-02  |
| 13         | 47    | 68.1        | 271    | 16 | R88823 | S. pyogenes DNaseB an  | 1.14e-02  |
| 14         | 47    | 68.1        | 271    | 11 | R58702 | Mitogenic factor asso  | 1.14e-02  |
| 15         | 47    | 68.1        | 293    | 12 | R70702 | DNA-ase-B.             | 1.14e-02  |
| 16         | 47    | 68.1        | 580    | 25 | W32431 | Mycobacterium tubercu  | 1.14e-02  |
| 17         | 47    | 68.1        | 580    | 25 | W32363 | Mycobacterium tubercu  | 1.14e-02  |
| 18         | 47    | 68.1        | 2183   | 7  | R39592 | L protein of attenuat  | 1.14e-02  |

| ID | Score | Query Match | Length | DB | ID     | Description           | Pred. No. |
|----|-------|-------------|--------|----|--------|-----------------------|-----------|
| 19 | 45    | 65.2        | 101    | 19 | W00728 | Vascular endothelial  | 1.88e-02  |
| 20 | 45    | 65.2        | 102    | 21 | W04824 | Vascular endothelial  | 1.88e-02  |
| 21 | 45    | 65.2        | 133    | 21 | W04828 | Vascular endothelial  | 1.88e-02  |
| 22 | 45    | 65.2        | 143    | 19 | W00727 | Vascular endothelial  | 1.88e-02  |
| 23 | 45    | 65.2        | 188    | 21 | W04829 | Fibrosarcoma vascular | 1.88e-02  |
| 24 | 45    | 65.2        | 188    | 19 | W00726 | Vascular endothelial  | 1.88e-02  |
| 25 | 45    | 65.2        | 188    | 21 | W04826 | Heart vascular endoth | 1.88e-02  |
| 26 | 45    | 65.2        | 188    | 19 | W00864 | Heart vascular endoth | 1.88e-02  |
| 27 | 45    | 65.2        | 195    | 21 | W04827 | Murine VRF167.        | 1.88e-02  |
| 28 | 45    | 65.2        | 207    | 19 | W00725 | Heart vascular endoth | 1.88e-02  |
| 29 | 45    | 65.2        | 207    | 21 | W04831 | Vascular endothelial  | 1.88e-02  |
| 30 | 45    | 65.2        | 207    | 21 | W04830 | Vascular endothelial  | 1.88e-02  |
| 31 | 45    | 65.2        | 207    | 19 | W00863 | Murine VRF166.        | 1.88e-02  |
| 32 | 45    | 65.2        | 221    | 23 | W07611 | Human Vascular endoth | 1.88e-02  |
| 33 | 45    | 65.2        | 332    | 23 | W19737 | Sugar biosynthesis en | 1.88e-02  |
| 34 | 45    | 65.2        | 360    | 24 | W26319 | Herpesvirus salmistr  | 1.88e-02  |
| 35 | 45    | 65.2        | 360    | 11 | R55792 | Herpesvirus salmistr  | 1.88e-02  |
| 36 | 45    | 65.2        | 393    | 19 | W00934 | Human G-protein coupl | 1.88e-02  |
| 37 | 45    | 65.2        | 418    | 24 | W00934 | Quail Fil4 receptor t | 1.88e-02  |
| 38 | 45    | 65.2        | 956    | 6  | R32356 | Exfolory amino acid r | 1.88e-02  |
| 39 | 45    | 65.2        | 1068   | 9  | R43341 | P110.                 | 1.88e-02  |
| 40 | 45    | 65.2        | 1068   | 9  | R46294 | Ptdins 3-kinase 110 k | 1.88e-02  |
| 41 | 45    | 65.2        | 1068   | 8  | R43342 | Human p110.           | 1.88e-02  |
| 42 | 45    | 65.2        | 4544   | 9  | R47861 | Alpha 2-Macroglobulin | 1.88e-02  |
| 43 | 45    | 65.2        | 4544   | 11 | R60517 | Human alpha-2-MR.     | 1.88e-02  |
| 44 | 45    | 63.8        | 293    | 1  | P80845 | Sequence of Tritirach | 2.40e-02  |
| 45 | 44    | 63.8        | 462    | 8  | R39635 | Human 3-Hydroxy-3-met | 2.40e-02  |

## ALIGNMENTS

RESULT 1  
ID R89373 standard; peptide: 9 AA.  
AC R89373;  
DE 18-SEP-1996 (first entry)  
DT B7 naturally processed protein derived immunogenic peptide.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN W09603140-A1.  
PD 08-FEB-1996.  
PE 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI: 96-116784/12.  
PT Compen. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PT C1a1n 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g. the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 100.0%; Score 69; DB 18; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3.38e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 1 mprgvvtl 9  
OY 1 mprgvvtl 9

RESULT 2  
ID W46517 standard: Protein: 697 AA.  
AC W46517  
DT 23-JUN-1998 (first entry)  
DE Saccharomyces cerevisiae V1A viral capsid-polymerase fusion protein.  
KW Saccharomyces cerevisiae V1A viral capsid-polymerase fusion protein;  
SCV1A totivirus; SCV1A.  
OS Saccharomyces cerevisiae.  
FH Key Location/Qualifiers  
FT Domain 202..442  
FT /Label="Multimerisation domain"  
PN W09800525-A1.  
PD 08-JAN-1998.  
PE 26-JUN-1997: U11216.  
PR 02-JUL-1996: US-674351.  
PA (UTRY ) UNIV NEW YORK STATE RES FOUND.  
PI Bruenn JA, Yao W;  
DR WPI: 98-086952/08.  
DR N-PSDB: V05285.  
PT Viral capsid polypeptide capable of inhibiting viral packaging -  
PT comprises part of viral capsid proteins, useful in, e.g. recombinant  
PT protein production of totiviruses.  
PS Claim 9; Pages 43-44; 52pp; English.  
CC The present sequence represents a capsid-polymerase fusion protein  
CC of Saccharomyces cerevisiae virus Ia (SCV1A). The virus is a member of  
CC the totiviruses in which all the viral functions are encoded by a single  
CC double strand of DNA. SCV1A contains two overlapping open reading  
CC frames encoding a capsid protein and a polymerase protein (see V05285).  
CC The latter is produced by translational frameshifting. The capsid  
CC protein has been shown to inhibit viral packaging of its cognate virus  
CC in yeast. The invention relates to the use of the sequence encoding the  
CC capsid protein (amino acids 1-443 of this sequence) for conferring  
CC resistance in yeast to the SCV1A and SCV1A totiviruses which can infect  
CC yeast cultures especially those that are used to produce recombinant  
CC proteins. Expression of the capsid proteins in other hosts e.g. plants  
CC or animals, can also be used to inhibit virus packaging thereby  
CC preventing viral spread and further infection.  
SQ Sequence 697 AA.

Query Match 71.0%; Score 49; DB 29; Length 697;  
Best Local Similarity 55.6%; Pred. No. 6.92e+01;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 393 murgivd1 401  
OY 1 MPRGVVTL 9

RESULT 3  
ID R78167 standard: Protein: 1155 AA.  
AC R78167  
DT 28-DEC-1995 (first entry)  
DE Mouse alpha-d subunit.  
KW Beta-2 integrin alpha-d subunit; antinflammatory; arteriosclerosis;  
KW inflammatory bowel disease; asthma; knock-out mouse.  
OS Mus sp.  
PN W09517412-A1.  
PD 29-JUN-1995.  
PE 21-DEC-1994; U14832.  
PR 23-DEC-1993; US-173497.  
PR 05-AUG-1994; US-286889.  
PA (ICOS-) ICOS CORP.  
PI Gallatin WM, Van Der Vlieten M;  
DR WPI: 95-240603/31.  
DR N-PSDB: Q91713.  
PT Alpha sub-unit polypeptide of human beta 2 integrin - used to  
PT identify potential antinflammatory agents, for the treatment of  
PT graft arteriosclerosis, inflammatory bowel disease, asthma, etc.  
PS Disclosure: Page 118-123; 172pp; English.  
CC A probe based on human integrin alpha-d clone 19A2 (given in  
CC Q91712) was used to isolate mouse alpha-d cDNA clones from a thymic  
CC oligo-dt-primed library in lambda ZAP II. RACE PCR was used

CC to obtain a composite sequence (Q91713) encoding a putative  
CC mouse alpha-d clone.  
SQ Sequence 1155 AA.

Query Match 69.6%; Score 48; DB 14; Length 1155;  
Best Local Similarity 66.7%; Pred. No. 8.91e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 1 murgv111 9  
OY 1 MPRGVVTL 9

RESULT 4  
ID W23060 standard: Protein: 1155 AA.  
AC W23060  
DT 24-FEB-1998 (first entry)  
DE Mouse beta 2 integrin alpha d subunit.  
KW Beta 2 integrin alpha d subunit; mouse; cell migration;  
KW cell adhesion; phagocytosis; diabetes; arteriosclerosis;  
KW multiple sclerosis; asthma; psoriasis; lung inflammation;  
KW acute respiratory distress syndrome; rheumatoid arthritis;  
KW monoclonal antibody.  
OS Mus musculus.  
PN W09731099-A1.  
PD 28-AUG-1997.  
PE 24-FEB-1997: U02713.  
PR 22-FEB-1996; US-605672.  
PA (ICOS-) ICOS CORP.  
PI Gallatin WM, Van Der Vlieten M;  
DR WPI: 97-435154/40.  
DR N-PSDB: T79251.  
PT Hybridoma 199M and antibody secreted by it - specific for new rat  
PT beta2 integrin subunit, useful to detect subunit in cells and  
PT modulate its activity.  
PS Example 19; Page 150-155; 222pp; English.  
CC This polypeptide comprises a murine homologue of a novel human  
CC human beta 2 integrin alpha d subunit (see W23049). Its sequence  
CC was deduced from a composite cDNA clone (see T79251). Homology  
CC between the external domains of human and mouse alpha d is high,  
CC but between cytoplasmic domains is low, suggesting C-terminal  
CC functional differences. Recombinant alpha d polypeptides can be  
CC expressed in transformed host cells and used to raise antibodies  
CC or to assay for compounds that modulate alpha d activity. The  
CC cDNA clone is given in W23061.  
SQ Sequence 1155 AA.

Query Match 69.6%; Score 48; DB 26; Length 1155;  
Best Local Similarity 66.7%; Pred. No. 8.91e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 1 murgv111 9  
OY 1 MPRGVVTL 9

RESULT 5  
ID R78168 standard: Protein: 1161 AA.  
AC R78168  
DT 28-DEC-1995 (first entry)  
DE Mouse alpha-d subunit.  
KW Beta-2 integrin alpha-d subunit; antinflammatory; arteriosclerosis;  
KW inflammatory bowel disease; asthma; knock-out mouse.  
OS Mus sp.  
PN W09517412-A1.  
PD 29-JUN-1995.  
PE 21-DEC-1994; U14832.  
PR 23-DEC-1993; US-173497.  
PR 05-AUG-1994; US-286889.  
PA (ICOS-) ICOS CORP.  
PI Gallatin WM, Van Der Vlieten M;  
DR WPI: 95-240603/31.



PT Alpha sub-unit polypeptide of human beta 2 integrin - used to identify potential antiinflammatory agents, for the treatment of great arteriosclerosis, inflammatory bowel disease, asthma, etc.  
 PS. Claim 30: Page 118-123; 172pp: English.  
 CC Mouse cDNA encoding a putative homolog of human integrin alpha-d subunit was isolated from a mouse splenic random primed library and an oligo dt-printed cDNA library using 2 homologous alpha-d probes.  
 CC Sequence 1161 AA;  
 SQ

Query Match 69.6%; Score 48; DB 14; Length 1161;  
 Best Local Similarity 66.7%; Pred. No. 8.91e+01;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1 mrvgrv11 9  
 1 mrvgrv11 9  
 1 mrvgrv11 9

RESULT 6  
 ID W23061 standard; Protein; 1161 AA.  
 AC W23061.

DT 24-FEB-1998 (first entry)  
 DE Mouse beta 2 integrin alpha d subunit.  
 KW Beta 2 integrin alpha d subunit; mouse; cell migration; cell adhesion; phagocytosis; diabetes; atherosclerosis; multiple sclerosis; asthma; psoriasis; lung inflammation; acute respiratory distress syndrome; rheumatoid arthritis; hydridoma; monoclonal antibody.  
 KW Mus musculus.  
 OS W09731099-A1.

PN 28-AUG-1997.  
 PE 24-FEB-1997; U02713.  
 PR 22-FEB-1996; US-605672.  
 PA (ICOS-) ICOS CORP.

PI Galatin WM, Van Der Vliet M;  
 DR WPI; 97-435154/40.  
 DR N-PSDB; T79256.

PT Hydridoma 199W and antibody secreted by it - specific for new rat beta2 integrin subunit, useful to detect subunit in cells and modulate its activity  
 PS Example 20: Page 161-165; 222pp; English.  
 CC This polypeptide comprises the murine homologue of a novel human beta 2 integrin subunit, designated alpha d (see W23049). Its sequence was deduced from a cDNA clone (see T79256) isolated from spleen cDNA libraries. Alpha d is involved in cell migration, phagocytosis and cell-cell interaction. Recombinant alpha d polypeptides can be expressed in transformed host cells for use in assays for identifying antibodies or other compounds that modulate alpha d activity or which modulate the interaction between alpha d and a ligand, for treating or preventing diseases in which macrophages are implicated such as type I diabetes, atherosclerosis, multiple sclerosis, asthma, psoriasis, lung inflammation, acute respiratory distress syndrome and rheumatoid arthritis.  
 CC Sequence 1161 AA;  
 SQ

Query Match 69.6%; Score 48; DB 26; Length 1161;  
 Best Local Similarity 66.7%; Pred. No. 8.91e+01;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1 mrvgrv11 9  
 1 mrvgrv11 9  
 1 mrvgrv11 9

RESULT 7  
 ID W23428 standard; Protein; 97 AA.  
 AC W23428.

DT 08-JAN-1998 (first entry)  
 DE Mycobacterium tuberculosis antigen TbrA24.  
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant; skin testing; M. tuberculosis.  
 OS Mycobacterium tuberculosis.

PN W09709428-A2.  
 PD 13-MAR-1997.  
 PE 30-AUG-1996; U14674.  
 PR 12-JUL-1996; US-680574.  
 PR 01-SEP-1995; US-523436.  
 PR 22-SEP-1995; US-533634.  
 PR 22-MAR-1996; US-620874.  
 PR 05-JUN-1996; US-659683.  
 PA (CORI-) CORIXA CORP.

PI Campos-neo A, Dillon DC, Houghton R, Reed SG, Skeiky YAM;  
 PI Twardzik DR, Vedvick TH;  
 DR WPI; 97-192903/17.  
 DR N-PSDB; T91472.

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are useful in vaccines for prevention or treatment of tuberculosis, also for diagnosis  
 PS Example 3; Page 104-105; 168pp; English.  
 CC A new immunogenic polypeptide has been developed comprising an immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its variant differing only in conservative substitutions and/or modifications). The present sequence represents a M. tuberculosis antigen, TbrA24. The immunogenic protein, and fusion proteins containing one or more of the proteins or one of the proteins plus ESAT-6, are useful in vaccines, preferably when formulated with a non-specific adjuvant, to induce an immune response against M. tuberculosis (for treatment or prevention).  
 CC Sequence 97 AA;  
 SQ

Query Match 68.1%; Score 47; DB 25; Length 97;  
 Best Local Similarity 75.0%; Pred. No. 1.14e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 41 vpkgrvvt 48  
 1 mrvgrvvt 8  
 1 mrvgrvvt 8

RESULT 8  
 ID W23360 standard; Protein; 97 AA.  
 AC W23360.

DT 13-JAN-1998 (first entry)  
 DE Mycobacterium tuberculosis antigen TbrA24.  
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant; skin testing; M. tuberculosis.  
 OS Mycobacterium tuberculosis.  
 PN W09709429-A2.  
 PD 13-MAR-1997.  
 PE 30-AUG-1996; U14675.  
 PR 12-JUL-1996; US-680573.  
 PR 01-SEP-1995; US-523435.  
 PR 22-SEP-1995; US-532136.  
 PR 22-MAR-1996; US-620280.  
 PR 05-JUN-1996; US-658800.  
 PA (CORI-) CORIXA CORP.

PI Campos-neo A, Dillon DC, Houghton R, Reed SG, Skeiky YAM;  
 PI Twardzik DR, Vedvick TH;  
 DR WPI; 97-192904/17.  
 DR N-PSDB; T91409.

PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens - useful for diagnosis of M. tuberculosis infection  
 PS Example 3; Page 113; 190pp; English  
 CC A new immunogenic polypeptide has been developed comprising an immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its variant differing only in conservative substitutions and/or modifications). The present sequence represents a M. tuberculosis antigen, TbrA24. The immunogenic polypeptide can be used to diagnose M. tuberculosis infection by forming complexes with specific antibodies in the sample. Fragments of DNA encoding the immunogenic polypeptide can be used as diagnostic primers or probes and agents that bind to the antigen, especially monoclonal antibodies or equivalent polyclonal antibodies, are also used for diagnosis.  
 CC Sequence 97 AA;  
 SQ

Query Match 68.1%; Score 47; DB 25; Length 97;  
Best Local Similarity 75.0%; Pred. No. 1.14e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 41. vpkqvvvt 48  
11:11111  
QY 1 MPRGVVVT 8

## RESULT 9

ID M13653 standard; Protein; 151 AA.  
AC M13653;  
DT 30-JUL-1997 (first entry)  
DE Herpesvirus Saimiri ORF13.  
KM CTLA: endothelial cell; haematopoietic cell; cytokine; IL.  
OS Herpesvirus saimiri.  
PN W09704097-A2.  
PD 06-FEB-1997.  
PF 18-JUL-1996; U11889.  
PR 11-AUG-1995; US-514014.  
PS 19-JUL-1995; US-504032.  
PA (CEMV) GENETICS INST INC.  
PI Carlin M, Giannotti J, Golden-fleet M, Goldman S;  
PI Jacobs K, Kelleher K, Mi S, Neben S, Pittman D;  
DR WPI; 97-132638/12.  
N-PSDB: T61415.  
PT New nucleic acid encoding the CTLA-8 protein - modulates growth of  
PT vascular endothelial and haematopoietic cells and induces cytokine  
PT expression, for treating infection, auto-immune disease, etc.  
PS Claim 19; Page 36; 50pp; English.  
CC Example 1 describes the isolation of human CTLA-8 cDNA (T61413).  
CC The primers given in T61416 to T61418 were used. The region from  
CC amino acid 29 to amino acid 163 shows marked homology to portions of  
CC rat CTLA-8 (T61414), amino acids 18 to 150) and herpesvirus  
CC Saimiri ORF13 (T61415; amino acids 19 to 151).  
SQ Sequence 151 AA.

Query Match 68.1%; Score 47; DB 24; Length 151;  
Best Local Similarity 75.0%; Pred. No. 1.14e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 40. prsymvvtl 47  
11:11111  
QY 2 PRGVVVTL 9

RESULT 10  
ID R76571 standard; Protein; 151 AA.  
AC R76571;  
DT 05-DEC-1995 (first entry)  
DE Herpesvirus ORF13 product.  
KM CTLA-8; cytokine-like protein; inflammation mediator.  
OS Herpesvirus saimiri.  
PN W09518826-A.  
PD 13-JUL-1995.  
PF 03-JAN-1995; U00001.  
PR 05-JAN-1994; US-177747.  
PS 27-MAY-1994; US-250846.  
PA (INRM) INST NAT SANTE & RECH MEDICALE.  
PI (SCHE) SCHERING CORP.  
PI Bancheureau J, Djossou O, Fossiez F, Golstein P;  
PI Lebecque SE, Rouvier E;  
DR WPI; 95-255038/33.  
N-PSDB: Q92882.  
PT New protein CTLA-B that induces secretion of inflammatory mediators  
PT related nucleic acid and antibodies, useful for modulating physiology  
PT and development of cells.  
PS Claim 4; Page 54; 64pp; English.  
CC A human CTLA-8 gene was isolated (Q92883) that showed 66.4%  
CC identity to the putative protein (R76571) encoded by ORF13  
CC (Q92882) of herpesvirus Saimiri.  
SQ Sequence 151 AA;

Query Match 68.1%; Score 47; DB 14; Length 151;  
Best Local Similarity 75.0%; Pred. No. 1.14e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 40. prsymvvtl 47  
11:11111  
QY 2 PRGVVVTL 9

RESULT 11  
ID W02387 standard; Protein; 151 AA.  
AC W02387;  
DT 05-DEC-1996 (first entry)  
DE HVS13 (viral homologue of interleukin-17).  
KM HVS13; cytokine; interleukin-17 receptor; IL-17R.  
OS Herpesvirus saimiri.  
PN W09629408-A1.  
PD 26-SEP-1996.  
PF 21-MAR-1996; U04018.  
PR 23-MAR-1995; US-410535.  
PS 07-AUG-1995; US-538765.  
PA (IMV) IMMUNEX CORP.  
PI Fanslow WC, Spriggs MK, Yao Z;  
PI WPI; 96-443184/44.  
DR DNA encoding interleukin-17 receptor - useful for regulating immune  
PT and inflammatory responses, or to suppress graft rejection  
PT Example 1; Page 36; 52pp; English.  
PS Herpesvirus Saimiri ORF13-encoded protein HVS13 (W02387), a viral  
CC homologue of interleukin-17 (IL-17), was used as a fusion partner  
CC in a fusion protein with the FC portion (W02305) of human IgG1.  
CC The fusion protein was used to screen cells for expression of IL-17  
CC receptor. Murine T-cell thymoma EL4 cells were found to bind the  
CC HVS13/FC fusion, and were used as a source of cDNA (see also  
CC T33800) for the murine IL-17 receptor (W04184).  
SQ Sequence 151 AA;

Query Match 68.1%; Score 47; DB 19; Length 151;  
Best Local Similarity 75.0%; Pred. No. 1.14e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 40. prsymvvtl 47  
11:11111  
QY 2 PRGVVVTL 9

RESULT 12  
ID R70701 standard; protein; 229 AA.  
AC R70701;  
DT 14-AUG-1995 (first entry)  
DE Recombinant DNA-ase-B.  
KM DNA-ase-B; diagnostic; vaccine; cystic fibrosis therapy.  
OS Streptococcus pyogenes.  
PN W09500650-A.  
PD 05-JAN-1995.  
PF 18-MAY-1994; U05626.  
PR 23-JUN-1993; US-082845.  
PA (BECT) BECKMAN INSTR INC.  
PI Adams CW, Belet CM, Pang PP;  
DR WPI; 95-052087/07.  
PT New DNA encoding streptococcus pyogenes DNase B - for diagnosing  
PT S. pyogenes infection, also new promoter for expressing other  
PT proteins  
PS Claim 1; Fig 4; 97pp; English.  
CC This is the full-length sequence of Streptococcus pyogenes  
CC DNA-ase-B, which is a marker of S. pyogenes infection. The protein  
CC may be used as a diagnostic agent or vaccine for S. pyogenes, or  
CC in cystic fibrosis.  
SQ Sequence 229 AA;

Query Match 68.1%; Score 47; DB 12; Length 229;  
Best Local Similarity 55.6%; Pred. No. 1.14e+02;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 189 ipravvsm 197  
 QY 1 MPRGVVTL 9

RESULT 13  
 ID R8823 standard; Protein: 271 AA.  
 AC R8823:  
 DT 25-JUN-1996 (first entry)  
 DE S. pyogenes DNaseB and leader sequence.  
 KW DNase B; nuclease; cystic fibrosis; vaccine; immunoassay;  
 OS Streptococcus pyogenes strain ATCC 14289.  
 FH Key Location/Qualifiers  
 FT peptide 1..43  
 FT /label= sig\_peptide  
 FT 44..271  
 FT /label= Mat\_protein  
 PN MO9606174-A1.  
 PD 29-FEB-1996  
 PF 18-AUG-1994; U09450.  
 PR 18-AUG-1994; WO-009450.  
 PA (BEC1) BECKMAN INSTR INC.  
 PI Adams CW, Belet MC, Pang PPY;  
 DR WPI: 96-151377/15.  
 N-PSDB: T12774.  
 PT New DNA encoding Streptococcus pyogenes DNase B - for recombinant  
 prod. of the enzyme in other bacteria, useful in immunoassays or  
 for treating cystic fibrosis.  
 PS Claim 1: Page 67-70; 115pp; English.  
 CC Streptococcus pyogenes DNase B, including the leader peptide,  
 has the amino acid sequence given in R8821. The enzyme can  
 be obt'd. on a large scale by expression of encoding DNA (T12774)  
 CC in transformed host cells, esp. Escherichia coli. Inclusion  
 CC of the leader peptide facilitates purification of the recombinant  
 CC enzyme. The DNase B is useful in immunoassays to detect  
 CC anti-DNase B antibodies in serum as a marker for S. pyogenes  
 CC infection, and is also useful as a vaccine or for treatment, via  
 CC aerosol delivery, of cystic fibrosis.  
 SQ Sequence 271 AA;

Query Match 68.1%; Score 47; DB 16; Length 271;  
 Best Local Similarity 55.6%; Pred. No. 1.14e+02;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 231 ipravvsm 239  
 QY 1 MPRGVVTL 9

RESULT 14  
 ID R58702 standard; Protein: 271 AA.  
 AC R58702:  
 DT 29-MAR-1995 (first entry)  
 DE Mitogenic factor associated with group A Streptococci.  
 KW Mitogenic factor; microdetection; group A streptococci; spe;  
 KW erythrogenic toxin; streptococcal pyrogenic exotoxin; blastogens;  
 KW scarlet fever toxin; erythematous skin reaction; infectious disease;  
 OS Streptococcus pyogenes.  
 PN EP-613947-A.  
 PD 07-SRP-1994.  
 PF 31-JAN-1994; 101386.  
 PR 01-FEB-1993; JP-037383.  
 PA (SHIO) SHIONOGI & CO LTD.  
 PI Hara A, Hinuma Y, Igarashi H, Iwasaki M, Kishishita M;  
 PI Okumura K, Takeda Y, Yutsudo T;  
 DR WPI: 94-272994/34.  
 N-PSDB: Q71612.  
 PT New mitogenic factor gene from Streptococcus pyogenes - used to  
 PT develop prods. for the early diagnosis of infectious disease  
 PT caused by gp A streptococci

PS Claim 7: Page 12-13; 20pp; English.  
 CC R58702 shows a mitogenic factor which exhibits rabbit peripheral  
 CC blood lymphocyte mitogenicity and/or DNA hydrolysing activity. It is  
 CC strongly associated with group A Streptococci and the nucleotide  
 CC sequences can be used for the microdetection of the gene and provide  
 CC an early diagnosis of infectious disease caused by the bacteria.  
 CC (See also Q71613-26).  
 SQ Sequence 271 AA;

Query Match 68.1%; Score 47; DB 11; Length 271;  
 Best Local Similarity 55.6%; Pred. No. 1.14e+02;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 231 ipravvsm 239  
 QY 1 MPRGVVTL 9

RESULT 15  
 ID R70702 standard; Protein: 293 AA.  
 AC R70702;  
 DT 15-AUG-1995 (first entry)  
 DE DNA-ase-B.  
 KW DNA-ase-B; diagnostic; vaccine; cystic fibrosis therapy.  
 OS Streptococcus pyogenes (ATCC 14289).  
 FH Key Location/Qualifiers  
 FT protein 46..293  
 FT /note= "mature protein"  
 FT peptide 1..45  
 FT /note= "leader peptide: claim 11"  
 FT misc\_difference 274 /note= "in-frame stop codon"  
 PN WO9500650-A.  
 PD 05-JAN-1995.  
 PF 18-MAY-1994; U05626.  
 PR 23-JUN-1993; US-082845.  
 PA (BEC1) BECKMAN INSTR INC.  
 PI Adams CW, Belet MC, Pang PPY;  
 DR WPI: 95-052087/07.  
 N-PSDB: Q85037.  
 PT New DNA encoding Streptococcus pyogenes DNase B - for diagnosing  
 PT S. pyogenes infection, also new promoter for expressing other  
 PT proteins  
 PS Disclosure: Fig 5; 97pp; English.  
 CC The sequence shows a gene product corresponding to a Streptococcus  
 CC pyogenes DNA-ase-B. The protein is useful as a diagnostic agent,  
 CC vaccine or as an aerosol to treat excessive lung viscosity, e.g. in  
 CC cystic fibrosis.  
 SQ Sequence 293 AA;

Query Match 68.1%; Score 47; DB 12; Length 293;  
 Best Local Similarity 55.6%; Pred. No. 1.14e+02;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 233 ipravvsm 241  
 QY 1 MPRGVVTL 9

Search completed: Fri Sep 11 13:21:08 1998  
 Job time : 16 secs.

**THIS PAGE BLANK (USPTO)**

(TM)

Distribution rights by Oxford Molecular Ltd

t generated.

1 MPRGVVTL 9

Gap 15

120441 seqs, 365

Listing first 45

1:plr1 2:plr2 3:plr3 4:plr4 5:nrl3d

Mean 23.179; Variance 30.626; scale 0.757

ved by analysis of the total score distribution.

## SUMMARIES

| Entry | Accession | Length | DB     | Description            | Pred. No. |
|-------|-----------|--------|--------|------------------------|-----------|
| 3.9.7 | 510       | 2      | A34160 | laurate omega-hydroxy  | 1.04e+00  |
| 3.5.4 | 181       | 1      | R5RT5  | ribosomal protein L5   | 4.11e+00  |
| 3.3.9 | 200       | 2      | S77490 | ribosomal protein L5   | 6.44e+00  |
| 3.3.9 | 511       | 2      | B34160 | cytochrome P450 4A7 -  | 6.44e+00  |
| 3.1.0 | 446       | 2      | X77743 | 4-aminobutyrate trans  | 1.00e+01  |
| 3.1.0 | 96        | 1      | SLX0A  | collipase A precursor  | 1.55e+01  |
| 3.1.0 | 96        | 1      | SLX0B  | collipase B precursor  | 1.55e+01  |
| 3.1.0 | 362       | 2      | A4601  | Na+/taurocholate tran  | 1.55e+01  |
| 3.1.0 | 697       | 2      | S50258 | capsid protein - Sacc  | 1.55e+01  |
| 3.1.0 | 831       | 2      | S50163 | nitrate reductase (un  | 1.55e+01  |
| 3.1.0 | 1154      | 2      | S43277 | hypothetical protein   | 1.55e+01  |
| 3.1.0 | 1154      | 2      | S43275 | hypothetical protein   | 1.55e+01  |
| 3.9.6 | 54        | 2      | Q0639  | hypothetical 5K prote  | 2.38e+01  |
| 3.9.6 | 246       | 2      | C64705 | gerC2 protein - Helic  | 2.38e+01  |
| 3.9.6 | 428       | 2      | U13871 | sensor kinase (EC 2.7  | 2.38e+01  |
| 3.9.6 | 538       | 2      | JC2457 | vascular cell adhesio  | 2.38e+01  |
| 3.9.6 | 552       | 1      | D2EP2  | phosphoribosylamino    | 2.38e+01  |
| 3.9.6 | 557       | 1      | D43322 | phosphoribosylamino    | 2.38e+01  |
| 3.9.6 | 571       | 1      | D2EYP  | phosphoribosylamino    | 2.38e+01  |
| 3.8.1 | 87        | 5      | 1EHND  | collipase, chain D - p | 3.64e+01  |
| 3.8.1 | 87        | 5      | 1EHC   | collipase, chain B - p | 3.64e+01  |
| 3.8.1 | 93        | 5      | 1PCB   | collipase B precursor  | 3.64e+01  |
| 3.8.1 | 93        | 5      | 1PCN   | collipase B precursor  | 3.64e+01  |

|    |    |      |       |   |                          |          |
|----|----|------|-------|---|--------------------------|----------|
| 24 | 47 | 66.1 | 95    | 2 | pancreatic colipase -    | 3.64e+01 |
| 25 | 47 | 66.1 | XLP62 | 1 | colipase II precursor    | 3.64e+01 |
| 26 | 47 | 66.1 | XLHU  | 1 | colipase precursor -     | 3.64e+01 |
| 27 | 47 | 66.1 | 1511  | 1 | immediate-early prote    | 3.64e+01 |
| 28 | 47 | 66.1 | 271   | 2 | mitogenic factor, 23k    | 3.64e+01 |
| 29 | 47 | 66.1 | 334   | 2 | transposase - Yersinia   | 3.64e+01 |
| 30 | 47 | 66.1 | 491   | 2 | C24829                   | 3.64e+01 |
| 31 | 47 | 66.1 | 825   | 2 | H+-transporting ATP s    | 3.64e+01 |
| 32 | 47 | 66.1 | 1830  | 1 | rftb protein - Vibrio    | 3.64e+01 |
| 33 | 47 | 66.1 | 1853  | 1 | myosin V - chicken       | 3.64e+01 |
| 34 | 47 | 66.1 | 2183  | 1 | myosin heavy chain, d    | 3.64e+01 |
| 35 | 47 | 66.1 | 2183  | 1 | RNA-directed RNA poly    | 3.64e+01 |
| 36 | 47 | 66.1 | 2183  | 1 | RNA-directed RNA poly    | 3.64e+01 |
| 37 | 47 | 66.1 | 4543  | 2 | LDL receptor-related     | 3.64e+01 |
| 38 | 46 | 66.7 | 99    | 2 | hypothetical protein     | 5.53e+01 |
| 39 | 46 | 66.7 | 123   | 1 | galactin precursor - b   | 5.53e+01 |
| 40 | 46 | 66.7 | 249   | 2 | N-acetyl gamma-glutam    | 5.53e+01 |
| 41 | 46 | 66.7 | 340   | 2 | N-acetylglutamate-1-phos | 5.53e+01 |
| 42 | 46 | 66.7 | 382   | 1 | galactokinase (EC 2.7    | 5.53e+01 |
| 43 | 46 | 66.7 | 382   | 1 | galactokinase (EC 2.7    | 5.53e+01 |
| 44 | 46 | 66.7 | 509   | 2 | cytochrome P450 Cyp4a    | 5.53e+01 |
| 45 | 46 | 66.7 | 543   | 2 | phosphoribosylaminoam    | 5.53e+01 |
| 46 | 46 | 66.7 | 557   | 2 | phosphoribosylaminoam    | 5.53e+01 |

## ALIGNMENTS

|  |  |  |
|--|--|--|
| RESULT   | 1  |  |
| ENTRY  |  |  |
| TITLE  | A34160   | #type complete<br>laurate omega-hydroxylase (EC 1.14.15.3) cytochrome P450 4A6<br>- rabbit |
| ALTERNATE_NAMES  | cytochrome P450a-1; cytochrome P450LGA omega 1               |  |
| ORGANISM   | #format_name Oryctolagus cuniculus                           | #common_name domestic<br>rabbit  |
| DATE   | 31-Mar-1992  | #sequence-revision 31-Mar-1992 #text-change  |
| ACCESSIONS   | A34160   | 23-Jan-1998  |
| REFERENCES   | A34160   | B34260; P00047; S23949   |
| authors  | Yokotani, N.; Bernhardt, R.; Sogawa, K.; Kusunose, E.; Gotoh |  |
| #journal   | O.; Kusunose, M.; Fujii-Kuriyama, Y.                         |  |
| #title   | J. Biol. Chem. (1989) 264:21665-21669                        |  |
|  | Two forms of omega-hydroxylase toward prostaglandin A and    |  |
|  | laurate. cDNA cloning and their expression.                  |  |
| #cross-references MVID:9094341                                   |  |  |
| #accession   | A34160   |  |
| ##molecule_type  | mRNA   |  |
| ##residues   | 1-510  | #label YOK   |
| #cross-references GB:M29531; NID:g164986; PID:g164987; GB:J05150 |  |  |
| REFERENCE  |  |  |
| authors  | Johnson, E.F.; Walker, D.L.; Griffin, K.J.; Clark, J.E.;     |  |
| #journal   | Okita, R.T.; Muenhoff, A.S.; Masters, B.S.                   |  |
| #title   | Biochemistry. (1990) 29:873-879                              |  |
|  | Cloning and expression of three rabbit kidney cDNAs encoding |  |
|  | lauric acid omega-hydroxylases.                              |  |
| #cross-references MVID:90254128                                  |  |  |
| #accession   | B34260   |  |
| ##molecule_type  | mRNA   |  |
| ##residues   | 1-423, 'YW', 426-434, 'R', 436-475, 'V', 477-510             | #label JOH   |
| #cross-references GB:M28656                                      |  |  |
| REFERENCE  |  |  |
| authors  | Kikuta, Y.; Kusunose, E.; Okumoto, T.; Kubota, I.; Kusunose, |  |
| #journal   | M.   |  |
| #title   | J. Biochem. (1990) 107:280-286                               |  |
|  | Purification and characterization of two forms of cytochrome |  |
|  | P-450 with omega-hydroxylase activities toward               |  |
|  | prostaglandin A and fatty acids from rabbit liver            |  |
|  | microsomes.  |  |
| #cross-references MVID:90299866                                  |  |  |
| #accession   | P00047   |  |
| ##molecule_type  | protein  |  |
| ##residues   | 5-24   | #label KIK   |
| ##experimental_source  | liver  |  |
| ##note   | amino-terminal sequence                                      |  |
| REFERENCE  |  |  |
|  | S23949   |  |

#authors Muerhoff, A.S.; Griffin, K.J.; Johnson, E.F.  
 #journal Arch. Biochem. Biophys. (1992) 296:66-72  
 #title Characterization of a rabbit gene encoding a  
 clofibrate-inducible fatty acid omega-hydroxylase: CYP4A6.  
 #cross-references MUID:92296782  
 #accession S23949  
 #status Preliminary  
 #molecule-type DNA  
 #residues 1-26, 'C', 28-379, 'M', 381-423, 'W', 426-434, 'R', 436-475,  
 'V', 477-510 ##label MUE  
 #comment This enzyme catalyzes the omega-hydroxylation of prostaglandin A1  
 and A2, as well as the omega- and (omega-1)-hydroxylation of  
 fatty acid.

GENETICS  
 #gene CYP4A6  
 CLASSIFICATION #superfamily cytochrome P450  
 #keywords chromoprotein; electron transfer; heme; iron; monooxygenase;  
 oxidoreductase; transmembrane protein

FEATURE  
 457 #binding-site heme iron (Cys) (axial ligand) #status  
 predicted

SUMMARY  
 #length 510 #molecular-weight 58129 #checksum 6263  
 Query Match 79.7%; Score 55; DB 2; Length 510;  
 Best Local Similarity 66.7%; Pred. No. 1.04e+00;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 405 LKGVITL 413  
 QY 1 MPRGVVTL 9

RESULT 2  
 ENTRY R5K75 #type complete  
 TITLE ribosomal protein L5 - Cyanophora paradoxa cyanelle  
 ORGANISM #formal\_name cyanelle Cyanophora paradoxa  
 DATE 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change  
 05-Sep-1997  
 ACCESSIONS S07067; S12216  
 #authors Bryant, D.A.; Stirewalt, V.L.  
 #journal FEBS Lett. (1990) 259:273-280  
 #title The cyanelle genome of Cyanophora paradoxa encodes ribosomal  
 proteins not encoded by the chloroplast genomes of higher  
 plants.  
 #cross-references MUID:90092562  
 #accession S07067  
 ##molecule-type DNA  
 ##residues 1-181 #label BRV  
 ##cross-references EMBL:X16548; NID:g11287; PID:g11288  
 S12211  
 #authors Michalowski, C.B.; Pfanzagl, B.; Loeffelhardt, W.; Bohmert,  
 H.J.  
 #journal Mol. Gen. Genet. (1990) 224:222-231  
 #title The cyanelle S10 spc ribosomal protein gene operon from  
 Cyanophora paradoxa.  
 #cross-references MUID:91117189  
 #accession S12216  
 ##molecule-type DNA  
 ##residues 1-145, 'G', 147-163, 'D', 165-181 #label MIC  
 ##cross-references GB:M30487; NID:9336645; PID:9336651

GENETICS  
 #gene rpl5; rplE  
 #map\_position 47-48  
 #genome cyanelle  
 CLASSIFICATION #superfamily Escherichia coli ribosomal protein L5  
 #keywords cyanelle; protein biosynthesis; ribosome  
 SUMMARY #length 181 #molecular-weight 20482 #checksum 584

Query Match 75.4%; Score 52; DB 1; Length 181;  
 Best Local Similarity 77.8%; Pred. No. 4.11e+00;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 84 MPRGVVTL 92  
 QY 1 MPRGVVTL 9

RESULT 3  
 ENTRY S77490 #type complete  
 TITLE ribosomal protein L5 - Synecocystis sp. (PCC 6803)  
 ALTERNATE\_NAMES protein s11808  
 ORGANISM #formal\_name Synecocystis sp.  
 #variety PCC 6803  
 DATE 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change  
 30-Jan-1998

ACCESSIONS S77490  
 REFERENCE S74322  
 #authors Kaneo, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;  
 Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugitara, M.;  
 Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;  
 Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpō,  
 S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;  
 Yasuda, M.; Tabata, S.  
 #journal DNA Res. (1996) 3:109-136  
 #title Sequence analysis of the genome of the unicellular  
 cyanobacterium Synecocystis sp. PCC6803. II. Sequence  
 determination of the entire genome and assignment of  
 potential protein-coding regions.  
 #cross-references MUID:97061201  
 #accession S77490  
 ##status nucleic acid sequence not shown; translation not shown  
 ##molecule-type DNA  
 ##residues 1-200 #label KAN  
 ##cross-references EMBL:D90905; NID:g1652360; PID:d1018070; PID:g1652415  
 ##note the nucleotide sequence was submitted to the EMBL data  
 Library, June 1996

GENETICS  
 #gene rpl5  
 #start\_codon GTG  
 CLASSIFICATION #superfamily Escherichia coli ribosomal protein L5  
 #keywords protein biosynthesis; ribosome  
 SUMMARY #length 200 #molecular-weight 22508 #checksum 5351

Query Match 73.9%; Score 51; DB 2; Length 200;  
 Best Local Similarity 77.8%; Pred. No. 6.44e+00;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 104 MPRGVVTL 112  
 QY 1 MPRGVVTL 9

RESULT 4  
 ENTRY B34160 #type complete  
 TITLE cytochrome P450 4A7 - rabbit  
 ALTERNATE\_NAMES cytochrome P450Ka-2; cytochrome P450Kc  
 ORGANISM #formal\_name Oryctolagus cuniculus #common\_name domestic  
 rabbit  
 DATE 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change  
 23-Jan-1998

ACCESSIONS B34160; C34260; JN0090  
 REFERENCE A34160  
 #authors Yokotani, N.; Bernhardt, R.; Sogawa, K.; Kusunose, E.; Gotoh,  
 O.; Kusunose, M.; Fujii-Kuriyama, Y.  
 #journal J. Biol. Chem. (1989) 264:21665-21669  
 #title Two forms of omega-hydroxylase toward prostaglandin A and  
 laurate. cDNA cloning and their expression.  
 #cross-references MUID:90094341  
 #accession B34160  
 ##molecule-type mRNA  
 ##residues 1-511 #label YOK  
 ##cross-references GB:M29530; NID:g164984; PID:g164985; GB:J05150  
 A34260  
 #authors Johnson, E.F.; Walker, D.L.; Griffin, K.J.; Clark, J.E.;  
 Okita, R.T.; Muerhoff, A.S.; Masters, B.S.

```

#journal      Biochemistry (1990) 29:873-879
#title        Cloning and expression of three rabbit kidney cDNAs encoding
               lauric acid omega-hydroxylases.
#cross-references M01D:90254128
#accession    C34260
#molecule-type mRNA
#residues     1-98, 'C', 100-149, 'F', 151-391, 'SK', 394-476, 'V', 478-511
               ##label JOH
#cross-references GB:M28657; NID:g164978; PID:g164979
#journal      JN0083
#authors      Yoshimura, R.; Kusunose, E.; Yokotani, N.; Yamamoto, S.;
               Kubota, I.; Kusunose, M.
#journal      J. Biochem. (1990) 108:544-548
#title        Purification and characterization of two forms of fatty acid
               omega-hydroxylase cytochrome P-450 from rabbit kidney
               cortex microsomes.
#cross-references M01D:91154157
#accession    JN0090
#molecule-type protein
#residues     5-7, 'X', 9-15, 'X', 17-22, 'X', 24 ##label YOS
#experimental-source kidney
#comment       This protein catalyzes the omega- and (omega-1)-hydroxylation of
               fatty acids.
#classification #superfamily cytochrome P450
#keywords       chromoprotein; electron transfer; heme; iron; monooxygenase;
               oxidoreductase; transmembrane protein
#feature       458
               #binding-site heme iron (Cys) (axial ligand) #status
               predicted
SUMMARY        #length 511 #molecular-weight 58337 #checksum 4545
Query Match    73.9%; Score 51; DB 2; Length 511;
Best Local Similarity 44.4%; Pred. No. 6.44e+00;
Matches        4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
DB            406 LPRGIIITL 414
               1:1111111
OY            1 MPRGVVTL 9
RESULT         5
ENTRY          572743
TITLE          4-aminobutyrate transaminase (EC 2.6.1.19) gabT -
               Mycobacterium leprae
ALTERNATE_NAMES 4-aminobutyrate aminotransferase gabT; B1177_F2.67 protein
ORGANISM       19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
DATE           09-Sep-1997
ACCESSIONS     S72743
REFERENCE      S72693
#authors       Smith, D.R.; Robison, K.
#submission    Submitted to the EMBL Data Library, November 1993
#description    Mycobacterium leprae cosmid B1177.
#accession     S72743
#status        preliminary
#molecule-type DNA
#residues      1-446 ##label SMT
#cross-references EMBL:U00011; NID:9466807; PID:9466832
GENETICS
#gene          gabT
#start_codon   GTG
#keywords       aminotransferase
SUMMARY        #length 446 #molecular-weight 47215 #checksum 5414
Query Match    72.5%; Score 50; DB 2; Length 446;
Best Local Similarity 77.8%; Pred. No. 1.00e+01;
Matches        7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
DB            30 VPRGVVTL 38
               1:1111111
OY            1 MPRGVVTL 9

```

```

RESULT         6
ENTRY          X1HOA
TITLE          collipase A precursor - horse
ALTERNATE_NAMES #formal_name Equus caballus #common_name domestic horse
ORGANISM        14-Nov-1983 #sequence_revision 04-Dec-1986 #text_change
DATE            26-Apr-1996
ACCESSIONS      A03164; A91119; A90220
REFERENCE        A90652
#authors         Sternby, B.; Engstrom, A.; Hellman, U.; Vihert, A.M.;
               Sternby, N.H.; Borgstrom, B.
#journal         Biochim. Biophys. Acta (1984) 784:75-80
#title           The primary sequence of human pancreatic collipase.
#cross-references M01D:84104937
#accession       A03164
#molecule-type protein
#residues        1-96 ##label STE
#note            residues 56-59 were positioned by homology; no overlap
               was obtained for 65-66
REFERENCE        A91119
#authors         Pierrot, M.; Astier, J.P.; Astier, M.; Charles, M.; Dreuth,
               J.
#journal         Eur. J. Biochem. (1982) 123:347-354
#title           Pancreatic collipase: crystallographic and biochemical
               aspects.
#cross-references M01D:82186702
#accession       A91119
#molecule-type protein
#residues        1-88, 'N', 90-91, 'K', 93 ##label PIE
REFERENCE        A90220
#authors         Julien, R.; Bechis, G.; Gregoire, J.; Rachelot, J.; Rochat,
               H.; Sarda, L.
#journal         Biochem. Biophys. Res. Commun. (1980) 95:1245-1252
#title           Evidence for the existence of two isocollipases in horse
               pancreas.
#cross-references M01D:81021166
#accession       A90220
#molecule-type protein
#residues        1-21, 'Q', 23-55 ##label JUL
#comment         Collipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms
               a 1:1 stoichiometric complex with it, enabling it to hydrolyze
               its substrate at the lipid-water interface. Without collipase the
               enzyme is washed off by bile salts, which are known to have an
               inhibitory effect on the lipase.
#classification #superfamily collipase
#keywords        lipid digestion; lipid hydrolysis; pancreas
#feature         1-5
               #domain propeptide #status experimental #label PRO
               #product collipase A #status experimental #label MAT
               17-87, 23-39, 27-63, #disulfide_bonds #status predicted
               28-61, 49-69 #binding_site micellar substrate (Trp, Tyr, Tyr)
               52, 55, 58, 59 #status predicted
SUMMARY        #length 96 #molecular-weight 10488 #checksum 2704
Query Match    71.0%; Score 49; DB 1; Length 96;
Best Local Similarity 62.5%; Pred. No. 1.55e+01;
Matches        5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
DB            4 PRGVVTL 11
               1:1111111
OY            2 PRGVVTL 9
RESULT         7
ENTRY          X1HOB
TITLE          collipase B precursor - horse
ALTERNATE_NAMES #formal_name Equus caballus #common_name domestic horse
ORGANISM        14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change
DATE            26-Apr-1996
ACCESSIONS      A03165; B90220
REFERENCE        A90637

```

```

#authors      Bonicel, J.; Couchoud, P.; Foglizzo, E.; Desnuelle, P.;
#journal      Chapus, C.
#title        Biochim. Biophys. Acta (1981) 669:39-45
#cross-references MUID:82046794
#accession    A03165
#molecule-type protein
#residues     1-96 ##label BON
REFERENCE     A90220
#authors      Julien, R.; Bechis, G.; Gregoire, J.; Rathelot, J.; Rochat,
#journal      H.; Sarda, L.
#title        Biochem. Biophys. Res. Commun. (1980) 95:1245-1252
#cross-references MUID:81021166
#accession    B90220
#molecule-type protein
#residues     1-21,'E',23-28,'T',30-55 ##label JUL
COMMENT       Colipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms
a 1:1 stoichiometric complex with it, enabling it to hydrolyze
its substrate at the lipid-water interface. Without colipase the
enzyme is washed off by bile salts, which are known to have an
inhibitory effect on the lipase.
CLASSIFICATION #superfamily colipase
KEYWORDS      lipid digestion; lipid hydrolysis; pancreas
FEATURE
1-5           #domain propeptide #status experimental #label PRO
6-96          #product colipase B #status experimental #label MAT
17-87,23-39,27-63, #disulfide_bonds #status predicted
28-61,49-69      #binding_site micellar substrate (Trp, Tyr, Tyr, Tyr)
52,55,58,59      #status predicted
SUMMARY        #length 96 #molecular-weight 10491 #checksum 3464
Query Match    71.0%; Score 49; DB 1; Length 96;
Best Local Similarity 62.5%; Pred. No. 1.55e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 4 PRGVVTL 11
1 1111111
2 PRGVVTL 9
RESULT 8
ENTRY      A41601 #type complete
TITLE      Na+/taurocholate transport protein - rat
ORGANISM   #formal_name Rattus norvegicus #common_name Norway rat
DATE       30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
16-Feb-1997
ACCESSIONS A41601
REFERENCE   A41601
#authors    Hagenbuch, B.; Stieger, B.; Foguet, M.; Luebbert, H.; Meier,
#journal     Proc. Natl. Acad. Sci. U.S.A. (1991) 88:10629-10633
#title       Functional expression cloning and characterization of the
hepatocyte Na(+)/bile acid cotransport system.
#cross-references MUID:92073340
#accession  A41601
#status     Preliminary
#molecule-type mRNA
#residues   1-362 ##label HAG
#cross-references GB:M7429
KEYWORDS     transmembrane protein
SUMMARY      #length 362 #molecular-weight 39295 #checksum 8711
Query Match    71.0%; Score 49; DB 2; Length 362;
Best Local Similarity 55.6%; Pred. No. 1.55e+01;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 57 KPGVYVAL 65
1 1111111
2 KPGVYVAL 9
OY 1 MPRGVVTL 9
```

```

RESULT 9
ENTRY      S50258 #type complete
TITLE      Capsid protein - Saccharomyces cerevisiae virus L-A
ORGANISM   #formal_name Saccharomyces cerevisiae virus L-A, SCV-L-A
DATE       13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change
09-Sep-1997
ACCESSIONS S50258
REFERENCE   S50258
#authors    Park, C.; Lopinski, J.D.; Tzeng, T.; Bruenn, J.A.
#journal     Submitted to the EMBL Data Library, August 1993
#title       Close relationship between two double-stranded RNA viruses of
Saccharomyces cerevisiae.
#accession  S50258
#molecule-type DNA
#residues   1-697 ##label PAR
#cross-references EMBL:U01060; NID:9595249; PID:9595250
GENETICS
SUMMARY      #length 697 #molecular-weight 78315 #checksum 3673
Query Match    71.0%; Score 49; DB 2; Length 697;
Best Local Similarity 55.6%; Pred. No. 1.55e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 393 MNRGIYDL 401
1 1111111
2 MNRGIYDL 9
OY 1 MPRGVVTL 9
RESULT 10
ENTRY      S50163 #type complete
TITLE      nitrate reductase (unclassified) (EC 1.-.-.-) large chain
precursor, periplasmic - Rhizosphaera pantotropha
ORGANISM   #formal_name Rhizosphaera pantotropha
DATE       16-Feb-1995 #sequence_revision 26-Jul-1996 #text_change
12-Sep-1997
ACCESSIONS S50163; S56135; S56128
REFERENCE   S50160
#authors    Berks, B.C.; Richardson, D.J.; Reilly, A.; Willis, A.C.;
#journal     Ferguson, S.J.
#title       submitted to the EMBL Data Library, August 1994
#description The periplasmic nitrate reductase operon of Rhizosphaera
pantotropha.
#accession  S50163
#molecule-type DNA
#residues   1-831 ##label BER
#cross-references EMBL:Z36773; NID:9600089; PID:9600093
REFERENCE   S56128
#authors    Berks, B.C.; Richardson, D.J.; Reilly, A.; Willis, A.C.;
#journal     Ferguson, S.J.
#title       The napEDABC gene cluster encoding the periplasmic nitrate
reductase system of Rhizosphaera pantotropha.
#accession  S56135
#status     nucleic acid sequence not shown
#molecule-type DNA
#residues   1-31,42-89,154-203 ##label BEW
#cross-references EMBL:Z36773
#accession  S56128
#molecule-type protein
#residues   114-127,'D',129-130,139-157,317-339,451-469,591-600;
642-657;675-694;699-715 ##label BEF
GENETICS
SUMMARY      napA
#gene        blocked amino end; oxidoreductase
KEYWORDS     1-31
#domain signal sequence #status predicted #label SIG\
#product nitrate reductase large chain #status predicted
#label MAT
SUMMARY      #length 831 #molecular-weight 92617 #checksum 1537
Query Match    71.0%; Score 49; DB 2; Length 831;
```



Best Local Similarity 100.0%; Pred. No. 1.55e+01;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 788 MPRGV 763

OY 1 MPRGV 6

RESULT 11

ENTRY 11 #type complete  
TITLE hypothetical protein 2 - Neurospora crassa retrotransposon

ORGANISM #formal\_name Neurospora crassa

DATE 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Sep-1997

ACCESSIONS

REFERENCE S43277

#authors Cambareri, E.B.; Helber, J.; Kinsey, J.A.

#journal Mol. Gen. Genet. (1994) 242:658-665

#title Tadi-1, an active LINE-like element of Neurospora crassa.

#accession S43277

#status preliminary; nucleic acid sequence not shown;

#molecule\_type DNA

#residues 1-1154 #label CAM

#cross-references EMBL:L25663; NID:g409762; PID:g409764

#note the nucleotide sequence was submitted to the EMBL Data

SUMMARY #length 1154 #molecular\_weight 130470 #checksum 5717

Query Match 71.0%; Score 49; DB 2; Length 1154;

Best Local Similarity 55.6%; Pred. No. 1.55e+01;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 828 KPRGIVIGL 836

OY 1 MPRGVVTL 9

RESULT 12 #type complete

ENTRY 12 hypothetical protein 2 - Neurospora crassa retrotransposon

ORGANISM #formal\_name Neurospora crassa

DATE 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Sep-1997

ACCESSIONS

REFERENCE S43274

#authors Cambareri, E.B.; Helber, J.; Kinsey, J.A.

#journal Mol. Gen. Genet. (1994) 242:658-665

#title Tadi-1, an active LINE-like element of Neurospora crassa.

#accession S43275

#molecule\_type DNA

#residues 1-1154 #label CAM

#cross-references EMBL:L25662; NID:g409759; PID:g409761

SUMMARY #length 1154 #molecular\_weight 130398 #checksum 5771

Query Match 71.0%; Score 49; DB 2; Length 1154;

Best Local Similarity 55.6%; Pred. No. 1.55e+01;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 828 KPRGIVIGL 836

OY 1 MPRGVVTL 9

RESULT 13

ENTRY 13 #type complete

TITLE hypothetical 5K protein (hisd 5' region) - Streptomyces

ORGANISM #formal\_name Streptomyces coelicolor

DATE 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Sep-1997

ACCESSIONS

REFERENCE J00639

#authors Limauro, D.; Avitabile, A.; Cappellano, C.; Puglila, A.M.;

#journal Bruni, C.B.

#title Gene (1990) 90:31-41

#accession J00639

#cross-references M0ID:90337345

#molecule\_type DNA

#residues 1-54 #label LIM

#cross-references GB:M1628; NID:g153295; PID:g153299

#experimental\_source strain A3(2)

SUMMARY #length 54 #molecular\_weight 5666 #checksum 3200

Query Match 69.6%; Score 48; DB 2; Length 54;

Best Local Similarity 55.6%; Pred. No. 2.38e+01;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 27 MPRGLIVL 35

OY 1 MPRGVVTL 9

RESULT 14

ENTRY 14 #type complete

TITLE gerc2 protein - Helicobacter pylori (strain 26695)

ORGANISM #formal\_name Helicobacter pylori

DATE 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 10-Oct-1997

ACCESSIONS

REFERENCE C64705

#authors Tomb, J.F.; White, O.; Kervayage, A.R.; Clayton, R.A.;

Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,

H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,

J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;

Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.;

McKenny, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;

Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;

Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;

Fujii, C.; Bowman, C.; Wathey, L.; Wallin, E.; Hayes,

W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser,

C.M.; Venter, J.C.

#journal Nature (1997) 388:539-547

#title The complete genome sequence of the gastric pathogen

#accession C64705

#status preliminary; nucleic acid sequence not shown;

#molecule\_type DNA

#residues 1-246 #label TOM

#cross-references GB:AE000647; GB:AE000511; NID:g2314645; PID:g2314655;

SUMMARY #length 246 #molecular\_weight 27871 #checksum 8177

Query Match 69.6%; Score 48; DB 2; Length 246;

Best Local Similarity 66.7%; Pred. No. 2.38e+01;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 155 KPRGVVTL 163

OY 1 MPRGVVTL 9

RESULT 15

ENTRY 15 #type complete

TITLE sensor kinase (EC 2.7.3.-) kinc - Bacillus subtilis

ALTERNATE\_NAMES protein-histidine kinase sdb; sporulation initiation

ORGANISM #formal\_name Bacillus subtilis

DATE 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change  
 20-Feb-1998  
 ACCESSIONS 139871, 139926; H69648  
 REFERENCE  
 #authors Ledoux, J.R.; Grossman, A.D.  
 #journal J. Bacteriol. (1995) 177:166-175  
 #title Isolation and characterization of *kinc*, a gene that encodes a  
 sensor kinase homologous to the sporulation sensor kinases  
*kina* and *kinB* in *Bacillus subtilis*.  
 #cross-references MUID:95095963  
 #accession 139871  
 #status Preliminary; translated from GB/EMBL/DBJ  
 #molecule\_type DNA  
 #residues 1-428 #label RES  
 #cross-references GB:134803; NID:g514329; PID:g514330  
 REFERENCE 139925  
 #authors Kobayashi, K.; Shoji, K.; Shimizu, T.; Nakano, K.; Sato, T.;  
 #journal J. Bacteriol. (1995) 177:176-182  
 #title Analysis of a suppressor mutation *ssb* (*kinc*) of *surO*B20  
 (*spo0A*) mutation in *Bacillus subtilis* reveals that *kinc*  
 encodes a histidine protein kinase.  
 #cross-references MUID:95095964  
 #accession 139926  
 #status Preliminary; translated from GB/EMBL/DBJ  
 #molecule\_type DNA  
 #residues 1-428 #label RE2  
 #cross-references GB:D37798; NID:g520575; PID:g520577  
 REFERENCE A69580  
 #authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;  
 Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;  
 Bojolin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,  
 A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;  
 Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
 Choi, S.K.; Codani, J.J.; Conner, I.F.; Cummings, N.J.;  
 Daniel, R.A.; Denizot, F.; Devine, K.M.; Diesterhoft, A.;  
 Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;  
 Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,  
 M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallon, N.; Ghim,  
 S.; Glaser, P.; Goffeau, A.; Golligly, E.J.; Grandi, G.;  
 Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,  
 C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;  
 Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;  
 Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,  
 Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;  
 Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;  
 Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;  
 Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,  
 M.; Moesti, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,  
 M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,  
 V.; Pohl, T.M.; Portetle, D.; Porwolik, S.; Prescott,  
 A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;  
 Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;  
 Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;  
 Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;  
 Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Solito,  
 B.; Sorokin, A.; Taccioni, E.; Takagi, T.; Takahashi, H.;  
 Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;  
 Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;  
 Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;  
 Wandutt, R.; Wedler, E.; Wedler, H.; Wiltzenegger, T.;  
 Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,  
 K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;  
 Yoshikawa, H.; Danchin, A.  
 #journal Nature (1997) 390:249-256  
 #title The complete genome sequence of the Gram-positive bacterium  
*Bacillus subtilis*.  
 #accession H69648  
 #status nucleic acid sequence not shown; translation not shown  
 #molecule\_type DNA  
 #residues 1-428 #label KUN  
 #experimental\_source strain 168  
 GENETICS

#gene *kinc*; *mskA*; *ssb*  
 KEYWORDS phosphotransferase; sporulation  
 SUMMARY #length 428 #molecular\_weight 48846 #checksum 4853  
 Query Match 69.6%; Score 48; DB 2; length 428;  
 Best Local Similarity 66.7%; Pred. No. 2.38e+01;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 DB 340 MPKGVTI 348  
 11:1111:  
 QY 1 MPKGVTI 9  
 Search completed: Fri Sep 11 13:21:53 1998  
 Job time : 27 secs.

\*\*\*\*\*  
 WIRE RELEASE  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Sep 11 13:22:12 1998; Maspar time 2.61 seconds  
 Tabular output not generated. 86,362 Million cell updates/sec

Title: >US-08-452-843-12  
 Description: (1-9) from US08452843.pep  
 Perfect score: 69  
 Sequence: 1 MPRGVVTL 9

Scoring table: PAM 150  
 Gap 15  
 Searched: 69111 segs, 25083644 residues  
 Post-processing: Minimum Match 08  
 Listing first 45 summaries

Database: swiss-prot35  
 1:swiss1

Statistics: Mean 24.256; Variance 26.038; scale 0.932

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Match | Length | ID | Description                         | Pred. No. |
|------------|-------|-------|--------|----|-------------------------------------|-----------|
| 1          | 56    | 81.2  | 597    | 1  | MCM3_ENTHI DNA REPLICATION LICENS   | 9.76e-02  |
| 2          | 55    | 79.7  | 510    | 1  | CP46_RABIT CYTOCHROME P450 IVA6 (   | 1.70e-01  |
| 3          | 52    | 75.4  | 181    | 1  | RK5_CVAP CYANELLE 50S RIBOSOMAL     | 8.60e-01  |
| 4          | 51    | 73.9  | 200    | 1  | R55_SYNV3 50S RIBOSOMAL PROTEIN     | 1.46e+00  |
| 5          | 51    | 73.9  | 511    | 1  | CP47_RABIT CYTOCHROME P450 IVA7 (   | 1.46e+00  |
| 6          | 50    | 72.5  | 446    | 1  | GABT_MYCLE 4-AMINO BUTYRATE AMINO   | 2.45e+00  |
| 7          | 49    | 71.0  | 65     | 1  | TRER_RHISN PROBABLE CONJUGAL TRAN   | 4.08e+00  |
| 8          | 49    | 71.0  | 106    | 1  | COLA_HORSE COLICOLIPASE A PRECURSOR | 4.08e+00  |
| 9          | 49    | 71.0  | 107    | 1  | COL_RABIT COLICOLIPASE A PRECURSOR  | 4.08e+00  |
| 10         | 49    | 71.0  | 108    | 1  | COLB_HORSE PROCOLIPASE B PRECURSOR  | 4.08e+00  |
| 11         | 49    | 71.0  | 362    | 1  | NTCP_RAT SODIUM/BILE ACID COTRA     | 4.08e+00  |
| 12         | 49    | 71.0  | 572    | 1  | MOES_LYTV MOESIN                    | 4.08e+00  |
| 13         | 49    | 71.0  | 831    | 1  | NAPA_RHOSH PERIPLASMIC NITRATE RE   | 4.08e+00  |
| 14         | 49    | 71.0  | 831    | 1  | NAPA_PARDT PERIPLASMIC NITRATE RE   | 4.08e+00  |
| 15         | 48    | 69.6  | 54     | 1  | YH11_STRCO HYPOTHEICAL 5.6 KD PR    | 6.74e+00  |
| 16         | 48    | 69.6  | 428    | 1  | KINC_BACSU SPORULATION KINASE C (   | 6.74e+00  |
| 17         | 48    | 69.6  | 552    | 1  | PURE_SCHPO PHOSPHORIBOSYLAMINOIM    | 6.74e+00  |
| 18         | 48    | 69.6  | 557    | 1  | PURE_VTGAC PHOSPHORIBOSYLAMINOIM    | 6.74e+00  |
| 19         | 48    | 69.6  | 571    | 1  | PURE_TFAST PHOSPHORIBOSYLAMINOIM    | 6.74e+00  |
| 20         | 47    | 68.1  | 95     | 1  | COL2_PIG COLICOLIPASE II            | 1.10e+01  |
| 21         | 47    | 68.1  | 112    | 1  | COL_HUMAN COLICOLIPASE PRECURSOR    | 1.10e+01  |
| 22         | 47    | 68.1  | 151    | 1  | VG13_HSVSA IMMEDIATE EARLY GENE 1   | 1.10e+01  |
| 23         | 47    | 68.1  | 334    | 1  | TRAB_YEREN TRANSPOSASE FOR INSERT   | 1.10e+01  |

## ALIGNMENTS

| RESULT | ID | MCM3_ENTHI | STANDARD: | PRT: | 597 AA.                            |
|--------|----|------------|-----------|------|------------------------------------|
| 24     | 47 | 68.1       | 491       | 1    | ATPB_CHIRE ATP SYNTHASE BETA CHAI  |
| 25     | 47 | 68.1       | 1829      | 1    | MYSD_CHICK DILUTE MYOSIN HEAVY CH  |
| 26     | 47 | 68.1       | 1853      | 1    | MYSA_MOUSE DILUTE MYOSIN HEAVY CH  |
| 27     | 47 | 68.1       | 2183      | 1    | RRPL_MEASA RNA POLYMERASE BETA SU  |
| 28     | 47 | 68.1       | 2183      | 1    | RRPL_MEASA RNA POLYMERASE BETA SU  |
| 29     | 47 | 68.1       | 4543      | 1    | LRPI_CHICK LOW-DENSITY LIPOPROTEI  |
| 30     | 46 | 66.7       | 123       | 1    | GALA_BOVIN GALANIN PRECURSOR       |
| 31     | 46 | 66.7       | 249       | 1    | ARGC_STRCO N-ACETYL-GAMMA-GLUTAM   |
| 32     | 46 | 66.7       | 340       | 1    | ARGC_STRCO N-ACETYL-GAMMA-GLUTAM   |
| 33     | 46 | 66.7       | 380       | 1    | GALI_SALT GALACTORINASE (EC 2.7.   |
| 34     | 46 | 66.7       | 381       | 1    | GALI_ECOLI GALACTORINASE (EC 2.7.  |
| 35     | 46 | 66.7       | 543       | 1    | PUR6_PICME PHOSPHORIBOSYLAMINOIM   |
| 36     | 46 | 66.7       | 557       | 1    | PUR6_SCHOC PHOSPHORIBOSYLAMINOIM   |
| 37     | 46 | 66.7       | 568       | 1    | PUR6_CANAL PHOSPHORIBOSYLAMINOIM   |
| 38     | 45 | 65.2       | 236       | 1    | PIRE_TIRE ORNATE PHOSPHORIBOSYL    |
| 39     | 45 | 65.2       | 360       | 1    | CPBH_HSVSA COMPLEMENT CONTROL PRO  |
| 40     | 45 | 65.2       | 385       | 1    | YEHY_ECOLI HYPOTHEICAL ABC TRANS   |
| 41     | 45 | 65.2       | 412       | 1    | Y360_MYCPN HYPOTHEICAL PROTEIN M   |
| 42     | 45 | 65.2       | 722       | 1    | Y360_MYCPN HYPOTHEICAL PROTEIN M   |
| 43     | 45 | 65.2       | 887       | 1    | ENOVY_HUMAN ENOVY-COA HYDRATASE (E |
| 44     | 45 | 65.2       | 1068      | 1    | HMCH_RAT 3-HYDROXY-3-METHYLGUT     |
| 45     | 45 | 65.2       | 1068      | 1    | PILA_BOVIN PHOSPHATIDYLINOSITOL 3  |

RESULT 1  
 ID MCM3\_ENTHI STANDARD: PRT: 597 AA.  
 AC Q24849;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE DNA REPLICATION LICENSING FACTOR MCM3.  
 GN MCM3.  
 OS ENTAMOEBA HISTOLYTICA.  
 OC EUKARYOTA: PROTOZOA: SARCOMASTIGOPHORA: SARCODINA: RHIZOPODA: LOBOSA:  
 CC AMOEBIIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HM-1:IMSS;  
 RA GANGOPADHYAY S.S., LOHIA A.;  
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: ACT AS A FACTOR THAT LICENSE THE DNA FOR ONE AND ONLY ONE ROUND OF REPLICATION PER CELL CYCLE. REQUIRED FOR DNA REPLICATION AND CELL PROLIFERATION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE MCM FAMILY.  
 DR EMBL: X98048; E245890; -  
 DR PROSITE: PS00847; MCM\_1; 1.  
 DR PROSITE: PS30051; MCM\_2; 1.  
 KM DNA REPLICATION; ATP-BINDING; NUCLEAR PROTEIN;  
 FT DOMAIN 180 386 MCM.  
 NP BIND 229 236 ATP (POTENTIAL);  
 SQ SEQUENCE 597 AA: 66412 MW: 36730499 CRR32;  
 Query Match 81.2%; Score 56; DB 1; Length 597;  
 Best Local Similarity 66.7%; Pred. No. 9.76e-02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 DB 104 MPRSVIVL 112  
 QY 1 MPRGVVTL 9  
 RESULT 2  
 ID CP46\_RABIT STANDARD: PRT: 510 AA.  
 AC P14580;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)  
 DE CYTOCHROME P450 IVA6 (EC 1.14.15.3) (LAURIC ACID OMEGA-HYDROXYLASE) (P450-KA-1).

GN CYP4A6.  
OS ORYCTOLAGUS CUNICULUS (RABBIT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; LAGOMORPHA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-KIDNEY;  
RX MEDLINE; 90254128.  
RA JOHNSON E.F., WALKER D.L., GRIFFIN K.J., CLARK J.E., OKITA R.T.,  
RA MEURHOFF A.S., MASTERS B.S.,  
RL BIOCHEMISTRY 29:873-879(1990).  
RN [2]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 5-24.  
RC TISSUE-KIDNEY;  
RX MEDLINE; 90094341.  
RA YOKOKAWA N., BERNHARDT R., SOGANA K., KUSUNOSE E., GOTOH O.,  
RA KUSUNOSE M., FUJII-KURIYAMA T.,  
RL J. BIOL. CHEM. 264:21665-21669(1989).  
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLOATE  
MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
ACIDS, AND XENOBIOTICS.  
CC -1- FUNCTION: THE KIDNEY P-450 SYSTEM IS RATHER SPECIALIZED FOR THE  
OMEGA-HYDROXYLATION OF FATTY ACIDS. P450-KA1 AND P450-KA2 CATALYZE  
THE OMEGA- AND (OMEGA-1)-HYDROXYLATION OF VARIOUS FATTY ACIDS WITH  
NO DRUG-METABOLIZING ACTIVITY, AND HYDROXYLATE PROSTAGLANDINE A1  
AND A2 SOLELY AT THE OMEGA-POSITION.  
CC -1- CATALYTIC ACTIVITY: OCTANE + REDUCED RUBREDOXIN + O(2) = 1-OCTANOL  
+ OXIDIZED RUBREDOXIN + H(2)O.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.  
CC -1- TISSUE SPECIFICITY: LIVER; KIDNEY.  
CC -1- INDUCTION: P450 CAN BE INDUCED TO HIGH LEVELS IN LIVER AND OTHER  
TISSUES BY VARIOUS FOREIGN COMPOUNDS, INCLUDING DRUGS, PESTICIDES,  
AND CARCINOGENS.  
CC EMBL; M28656; G164977; -;  
DR EMBL; M29531; A316487; -;  
DR PIR; A34160; A34160.  
DR PIR; B34260; B34260.  
DR PROSITE; P500086; CYTOCHROME\_P450; 1.  
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME;  
KW MICROSOME.  
FT PROPER. 1 4  
FT CHAIN 5 510 CYTOCHROME P450 IVAB.  
FT BINDING 457 457 HEME.  
FT CONFLICT 424 425 VW -> CG (IN REF. 2).  
FT CONFLICT 434 434 F -> S (IN REF. 2).  
FT CONFLICT 476 476 V -> L (IN REF. 2).  
SQ SEQUENCE 510 AA; 58300 MW; E11495ED CRC32;

Query Match 79.7%; Score 55; DB 1; Length 510;  
Best Local Similarity 66.7%; Pred. No. 1.70e-01;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 405 LPRGVVTL 413  
OY 1 MPRGVVTL 9

RESULT 3  
ID RK5\_CYP4A STANDARD; PRT; 181 AA.  
AC P14807;  
DT 01-APR-1990 (REL. 14, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE CYANELLE 50S RIBOSOMAL PROTEIN L5.  
DE RPL5.  
GN CYANOPHORA PARADOXA.  
OG CYANELLE.  
OC EUKARYOTA; PLANTA; PHYCOPHYTA; GLAUCOPHYTA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LB555 / PRINGSHEIM;

RX MEDLINE; 90092562.  
RA BRYANT D.A., STIREWALT V.L.;  
RL REBS LEFT. 259:273-280(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-UTEX 5550;  
RX MEDLINE; 91117189.  
RA MICHALOWSKI C.B., PFANZAGL B., LOEFFELHARDT W., BOHNERT H.J.;  
RL MOL. GEN. GENET. 224:222-231(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LB555 / PRINGSHEIM;  
RA STIREWALT V.L., MICHALOWSKI C.B., LUFFELHARDT W., BOHNERT H.J.,  
RA BRYANT D.A.;  
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: THIS IS ONE OF 3 PROTEINS THAT MEDIANE THE ATTACHMENT OF  
THE 5S RNA INTO THE LARGE RIBOSOMAL SUBUNIT (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE L5P FAMILY OF RIBOSOMAL PROTEINS.  
DR EMBL; M30487; G336651; -;  
DR EMBL; X16548; G11288; -;  
DR EMBL; U30821; G1016136; -;  
DR PIR; S07067; RSKT5.  
DR PROSITE; P500358; RIBOSOMAL\_L5; 1.  
KW RIBOSOMAL PROTEIN; RNA-BINDING; CYANELLE.  
FT CONFLICT 146 146 G -> S (IN REF. 1).  
FT CONFLICT 164 164 D -> N (IN REF. 1).  
SQ SEQUENCE 181 AA; 20453 MW; A4F2AA28 CRC32;

Query Match 75.4%; Score 52; DB 1; Length 181;  
Best Local Similarity 77.8%; Pred. No. 8.60e-01;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 84 MPRGVVTL 92  
OY 1 MPRGVVTL 9

RESULT 4  
ID RL5\_SVNY3 STANDARD; PRT; 200 AA.  
AC P73308;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE 50S RIBOSOMAL PROTEIN L5.  
GN RPL5 OR RPL5 OR S11808.  
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).  
OC PROKARYOTA; GRACILICUTES; OXYPHOTOBACTERIA;  
OC CYANOBACTERIA (BLUE-GREEN ALGAE); CHROOCOCCELES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97061201.  
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
RA MIYAIKAWA N., HIROSAWA M., SOGURA M., SASAMOTO S., KIMURA T.,  
RA HOSOGUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARDO K.,  
RA OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A.,  
RA YAMADA M., YASUDA M., TABATA S.;  
RL DNA RES. 3:109-136(1996).  
CC -1- FUNCTION: THIS IS ONE OF 3 PROTEINS THAT MEDIANE THE ATTACHMENT OF  
THE 5S RNA INTO THE LARGE RIBOSOMAL SUBUNIT (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE L5P FAMILY OF RIBOSOMAL PROTEINS.  
DR EMBL; D90905; G1652415; -;  
DR PROSITE; P500358; RIBOSOMAL\_L5; 1.  
KW RIBOSOMAL PROTEIN; RNA-BINDING.  
SQ SEQUENCE 200 AA; 22508 MW; E411C315 CRC32;

Query Match 73.9%; Score 51; DB 1; Length 200;  
Best Local Similarity 77.8%; Pred. No. 1.46e-00;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 104 MPRGVVTL 112  
OY 1 MPRGVVTL 9

RESULT 5  
ID CP47 RABIT STANDARD: PRT: 511 AA.  
AC P14581:  
DT 01-JAN-1990 (REL. 13, CREATED)  
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
DE CYTOCHROME P450 IVA7 (EC 1.14.15.3) (LAURIC ACID OMEGA-HYDROXYLASE)  
DE (P450-KA-2).  
GN CYP4A7  
OS ORYCTOLAGUS CUNICULUS (RABBIT).  
OC EUMARYPTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; LAGOMORPHA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-KIDNEY;  
RX MEDLINE: 90234128.  
RA JOHNSON E.F., WALKER D.L., GRIFFIN K.J., CLARK J.E., OKITA R.T.,  
RA MEURHOFF A.S., MASTERS B.S.;  
RA BIOCHEMISTRY 29:873-879(1990).  
RL [2]  
RN  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 5-24.  
RC TISSUE-KIDNEY;  
RX MEDLINE: 90094341.  
RA YOKOTANI N., BERNHARDT R., SOGAWA K., KUSUNOSE E., GOTOH O.,  
RA KUSUNOSE M., FUJII-KURIYAMA Y.;  
RL J. BIOL. CHEM. 264:21665-21669(1989).  
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
ACIDS, AND XENOBIOTICS.  
CC -1- FUNCTION: THE KIDNEY P-450 SYSTEM IS RATHER SPECIALIZED FOR THE  
OMEGA-HYDROXYLATION OF FATTY ACIDS. P450-KA1 AND P450-KA2 CATALYZE  
THE OMEGA- AND (OMEGA-1)-HYDROXYLATION OF VARIOUS FATTY ACIDS WITH  
NO DRUG-METABOLIZING ACTIVITY, AND HYDROXYLATE PROSTAGLANDINE A1  
AND A2 SOLELY AT THE OMEGA-POSITION.  
CC -1- CATALYTIC ACTIVITY: OCTANE + REDUCED RUBREDOXIN + O(2) = 1-OCTANOL  
+ OXIDIZED RUBREDOXIN + H(2)O.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.  
CC -1- TISSUE SPECIFICITY: LIVER, KIDNEY, SMALL INTESTINE.  
CC -1- INDUCTION: P450 CAN BE INDUCED TO HIGH LEVELS IN LIVER AND OTHER  
TISSUES BY VARIOUS FOREIGN COMPOUNDS, INCLUDING DRUGS, PESTICIDES,  
AND CARCINOGENS.  
DR EMBL: M28657; G164979; -;  
DR EMBL: M29350; G164985; -;  
DR PIR: B34160; C34260.  
DR PIR: B34160; B34160.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME;  
KW MICROSOME.  
FT PROPEP 1 4  
FT CHAIN 5 511  
FT BINDING 458 458  
FT CONFLICT 99 99  
FT CONFLICT 150 150  
FT CONFLICT 392 393  
FT CONFLICT 477 477  
FT CONFLICT 511 AA: 58318 MW; 2EA3C45E CRC32;  
SQ SEQUENCE

Query Match 73.9%; Score 51; DB 1; Length 511;  
Best Local Similarity 44.4%; Pred. No. 146+00;  
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DB 406 LPKGIITL 414  
QY 1 MPRGVVTL 9

RESULT 6  
ID GABT\_MYCLE STANDARD: PRT: 446 AA.  
AC P40829;  
DT 01-FEB-1995 (REL. 31, CREATED)

DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE 4-AMINOBUTYRATE AMINOTRANSFERASE (EC 2.6.1.19) (GAMMA-AMINO-N-BUTYRATE  
DE TRANSAMINASE) (GABA TRANSAMINASE) (GLUTAMATE:SUCCINIC SEMIALDEHYDE  
DE TRANSAMINASE) (GABA AMINOTRANSFERASE).  
GN GABT OR B1177\_F2-67.  
OS MYCOBACTERIUM LEPRAE.  
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SMITH D.R., ROBISON K.;  
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- CATALYTIC ACTIVITY: 4-AMINOBUTANOATE + 2-OSOGUTARATE = SUCCINATE  
SEMIALDEHYDE + L-GLUTAMATE.  
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -1- PATHWAY: 4-AMINOBUTYRATE (GABA) DEGRADATION PATHWAY.  
CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
AMINOTRANSFERASES.  
DR EMBL: U00011; G466832; -;  
DR PROSITE: PS00600; AA-TRANSFER\_CLASS\_3; 1.  
KW TRANSFERASE; AMINOTRANSFERASE; PYRIDOXAL PHOSPHATE.  
FT BINDING 291 291  
FT BINDING 291 291  
SQ SEQUENCE 446 AA: 47215 MW; 7F34F294 CRC32;

Query Match 72.5%; Score 50; DB 1; Length 446;  
Best Local Similarity 77.8%; Pred. No. 245+00;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 30 VPRGVVTL 38  
QY 1 MPRGVVTL 9

RESULT 7  
ID TRBK\_RHISN STANDARD: PRT: 65 AA.  
AC P55401;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE PROBABLE CONJUGAL TRANSFER PROTEIN TRBK PRECURSOR.  
GN TRBK OR Y4DB.  
OS RHIZOBIUM SP. (STRAIN NGR234).  
OC PLASMID SYN PGR234A.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC RHIZOBIACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 97305956.  
RA FREIBERG C.A., FELLAY R., BAIROCH A., BROUGHTON W.J., ROSENTHAL A.,  
RA PERRET X.;  
RL NATURE 387:394-401(1997).  
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).  
CC -1- SIMILARITY: STRONG, TO A. TUMEFACIENS TI PLASMID TRBK.  
DR EMBL: AE000068; G2182344; -;  
CC CONJUGATION; PERIPLASMIC; PLASMID; SIGNAL.  
FT SIGNAL 1 19  
FT CHAIN 20 65  
FT CHAIN 20 65  
SQ SEQUENCE 65 AA: 6864 MW; 909C5F10 CRC32;

Query Match 71.0%; Score 49; DB 1; Length 65;  
Best Local Similarity 44.4%; Pred. No. 408+00;  
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DB 1 MSRAVITL 9  
QY 1 MPRGVVTL 9

RESULT 8  
ID COLA\_HORSE STANDARD: PRT: 106 AA.  
AC P02704;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE PROCOLIPASE A PRECURSOR (FRAGMENT).  
 OS EQUUS CABALLUS (HORSE).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; PERISSODACTYLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PANCREAS;  
 RX MEDLINE: 94325330.  
 RA CRENON I., GRANON S., CHAPUS C., KERPELEC B.;  
 RL BIOCHIM. BIOPHYS. ACTA 1213:357-360(1994).  
 RN [2]  
 RP SEQUENCE OF 12-106.  
 RX MEDLINE: 84104937.  
 RA STERNBY B., ENGSTRÖM A., HELLMAN U., VIERHART A.M., STERNBY N.H.,  
 RL BIOCHIM. BIOPHYS. ACTA 784:75-80(1984).  
 RN [3]  
 RP SEQUENCE OF 12-106.  
 RX MEDLINE: 82186702.  
 RA PIERROT M., ASTIER J.-P., ASTIER M., CHARLES M., DRENTH J.;  
 RL EUR. J. BIOCHEM. 123:347-354(1982).  
 RN [4]  
 RP SEQUENCE OF 12-66.  
 RX MEDLINE: 81021166.  
 RA JULIEN R., BECHIS G., GREGOIRE J., RATHÉLOT J., ROCHAT H., SARDA L.;  
 RL BIOCHIM. BIOPHYS. RES. COMMUN. 95:1245-1252(1980).  
 CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS  
 THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT  
 COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN  
 INHIBITORY EFFECT ON THE LIPASE.  
 CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY  
 SIGNAL.  
 CC -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.  
 DR EMBL: X74503; G572677; -;  
 DR PIR: A03164; XLH0A.  
 DR HSSP: P02703; 1PCN.  
 DR PROSITE: PS00121; COLIPASE; 1.  
 KW LIPID DEGRADATION; DIGESTION; PANCREAS; SIGNAL.  
 FT NON TER 1 1  
 FT SIGNAL <1 1  
 FT PROPEP 12 16 ENTEROSTATIN, ACTIVATION PEPTIDE.  
 FT CHAIN 17 106 COLIPASE A.  
 FT DISULFID 28 39 BY SIMILARITY.  
 FT DISULFID 34 50 BY SIMILARITY.  
 FT DISULFID 38 72 BY SIMILARITY.  
 FT DISULFID 60 80 BY SIMILARITY.  
 FT DISULFID 74 98 BY SIMILARITY.  
 FT BINDING 63 63 BILE SALT MICELLES.  
 FT CONFLICT 33 33 Q -> E (IN REF. 2 AND 3).  
 FT CONFLICT 43 43 S -> E (IN REF. 2, 3 AND 4).  
 FT CONFLICT 100 100 D -> N (IN REF. 3).  
 FT CONFLICT 103 103 R -> K (IN REF. 3).  
 FT CONFLICT 106 106 E -> ER (IN REF. 3).  
 SQ SEQUENCE 106 AA; 11388 MW; 9970DB63 CRC32;  
 Query Match 71.0%; Score 49; DB 1; Length 106;  
 Best Local Similarity 62.5%; Pred. No. 4.08e+00;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUKARYOTA; LAGOMORPHA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PANCREAS;  
 RX MEDLINE: 93345715.  
 RA COLWELL N.S., ALEMAN-GOMEZ J.A., SASSER T.L., KUMAR V.B.;  
 RL INT. J. BIOCHEM. 25:885-890(1993).  
 CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS  
 THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT  
 COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN  
 INHIBITORY EFFECT ON THE LIPASE.  
 CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY  
 SIGNAL (BY SIMILARITY).  
 CC -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.  
 DR EMBL: L06329; G164894; -;  
 DR PROSITE: PS00121; COLIPASE; 1.  
 KW LIPID DEGRADATION; DIGESTION; PANCREAS; SIGNAL.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT PROPEP 18 22 ENTEROSTATIN, ACTIVATION PEPTIDE  
 FT CHAIN 23 107 COLIPASE.  
 FT DISULFID 34 45 BY SIMILARITY.  
 FT DISULFID 40 56 BY SIMILARITY.  
 FT DISULFID 44 78 BY SIMILARITY.  
 FT DISULFID 66 86 BY SIMILARITY.  
 FT DISULFID 80 104 BY SIMILARITY.  
 SQ SEQUENCE 107 AA; 11271 MW; 1D6F7BCE CRC32;  
 Query Match 71.0%; Score 49; DB 1; Length 107;  
 Best Local Similarity 62.5%; Pred. No. 4.08e+00;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 21 PRGVIINL 28  
 Oy 2 PRGVVTL 9  
 RESULT 10  
 ID COLB\_HORSE STANDARD: PRT: 108 AA.  
 AC P02705;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DE PROCOLIPASE B PRECURSOR (FRAGMENT).  
 OS EQUUS CABALLUS (HORSE).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; PERISSODACTYLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PANCREAS;  
 RX MEDLINE: 94325330.  
 RA CRENON I., GRANON S., CHAPUS C., KERPELEC B.;  
 RL BIOCHIM. BIOPHYS. ACTA 1213:357-360(1994).  
 RN [2]  
 RP SEQUENCE OF 14-108.  
 RX MEDLINE: 82046794.  
 RA BONICEL J.J., COUCHOUD P.M., FOGILIZO E., DESNUELLE P., CHAPUS C.;  
 RL BIOCHIM. BIOPHYS. ACTA 669:39-45(1981).  
 RN [3]  
 RP SEQUENCE OF 14-68.  
 RX MEDLINE: 81021166.  
 RA JULIEN R., BECHIS G., GREGOIRE J., RATHÉLOT J., ROCHAT H., SARDA L.;  
 RL BIOCHIM. BIOPHYS. RES. COMMUN. 95:1245-1252(1980).  
 CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS  
 THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT  
 COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN  
 INHIBITORY EFFECT ON THE LIPASE.  
 CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY  
 SIGNAL.  
 CC -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.  
 DR EMBL: X74344; G572679; -;  
 DR PIR: A03165; XLH0B.

Sun Sep 19 10:55:19 1998

DR HSSP: P02703; IPCN.  
 DR PROSITE: PS00121; COLIPASE: 1.  
 KM LIPID DEGRADATION; DIGESTION; PANCREAS; SIGNAL.  
 FT NON\_TER 1 1  
 FT SIGNAL <1 13  
 FT PROPEP 14 18 ENTEROSTATIN, ACTIVATION PEPTIDE.  
 FT CHAIN 19 108 COLIPASE B.  
 FT BINDING 65 65 BILE SALT MICELLES.  
 FT DISULFID 30 41 BY SIMILARITY.  
 FT DISULFID 36 52 BY SIMILARITY.  
 FT DISULFID 40 74 BY SIMILARITY.  
 FT DISULFID 62 82 BY SIMILARITY.  
 FT DISULFID 76 100 BY SIMILARITY.  
 FT CONFLICT 35 35 O -> E (IN REF. 3).  
 FT CONFLICT 42 42 H -> T (IN REF. 3).  
 FT CONFLICT 108 108 E -> ER (IN REF. 2).  
 SQ SEQUENCE 108 AA; 11618 MW; 1A17861D CRC32;  
 Query Match 71.0%; Score 49; DB 1; Length 108;  
 Best Local Similarity 62.5%; Pred. No. 4.08e+00;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 17 PRGVITL 24  
 QY 2 PRGVVTL 9  
 RESULT 11  
 ID NTCP\_RAT STANDARD; PRT; 362 AA.  
 AC P26435;  
 DT 01-AUG-1992 (REL. 23, CREATED)  
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
 DE SODIUM/BILE ACID COTRANSPORTER (NA+)/BILE ACID COTRANSPORTER)  
 DE (NA+)/TAUROCHOLATE TRANSPORT PROTEIN (SODIUM/TAUROCHOLATE  
 DE COTRANSPORTING POLYPEPTIDE).  
 GN SLC10A1 OR NTCP.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 92073340.  
 RA HAGENBUCH B., STIEGER B., FOGUET M., LUEBBERT H., MEIER P.J.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:10629-10633(1991).  
 CC -1- FUNCTION: THE HEPATIC SODIUM/BILE ACID UPTAKE SYSTEM EXHIBITS  
 CC BROAD SUBSTRATE SPECIFICITY & TRANSPORTS VARIOUS NONBILE ACID  
 CC ORGANIC COMPOUNDS AS WELL. IT IS STRICTLY DEPENDENT ON THE  
 CC EXTRACELLULAR PRESENCE OF SODIUM.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: LIVER AND KIDNEY.  
 CC -1- SIMILARITY: BELONGS TO THE SODIUM/BILE ACID SYMPORTER FAMILY  
 CC (Sbf).  
 DR EMBL: M77479; G206854; -  
 DR PIR: A41601; A41601  
 KM TRANSMEMBRANE; TRANSPORT; SYMPORT; SODIUM TRANSPORT; GLYCOPROTEIN.  
 FT TRANSMEM 24 45 POTENTIAL.  
 FT TRANSMEM 60 80 POTENTIAL.  
 FT TRANSMEM 82 98 POTENTIAL.  
 FT TRANSMEM 158 178 POTENTIAL.  
 FT TRANSMEM 190 211 POTENTIAL.  
 FT TRANSMEM 228 244 POTENTIAL.  
 FT TRANSMEM 285 306 POTENTIAL.  
 FT CARBOHYD 5 5 POTENTIAL.  
 FT CARBOHYD 11 11 POTENTIAL.  
 FT CARBOHYD 103 103 POTENTIAL.  
 FT CARBOHYD 117 117 POTENTIAL.  
 FT CARBOHYD 271 271 POTENTIAL.  
 SQ SEQUENCE 362 AA; 39295 MW; 69E1D9DC CRC32;  
 Query Match 71.0%; Score 49; DB 1; Length 362;  
 Best Local Similarity 55.6%; Pred. No. 4.08e+00;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Db 57 KPGVIVAL 65  
 QY 1 MPRGVVT 9  
 RESULT 12  
 ID MOES\_LYTV STANDARD; PRT; 572 AA.  
 AC P52962;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE MOESIN.  
 OS LITICHINUS VARIEGATUS (SEA URCHIN).  
 OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;  
 OC EUECHINOIDEA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95256329.  
 RA BACHMAN E.S., MCCILLY D.R.;  
 RL J. CELL SCI. 108:161-171(1995).  
 CC -1- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL  
 CC STRUCTURES TO THE PLASMA MEMBRANE.  
 CC -1- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,  
 CC RADIXIN, AND TALIN.  
 CC -1- SIMILARITY: VERY STRONG TO EZRIN AND RADIXIN.  
 DR EMBL: U14180; G719272; -  
 DR PROSITE: PS00660; BAND 4.1; 1.  
 DR PROSITE: PS00661; BAND 4.1.2; 1.  
 DR PROSITE: PS00657; BAND 4.1.3; 1.  
 KM STRUCTURAL PROTEIN; CYTOSKELETON.  
 FT DOMAIN 58 224  
 SQ SEQUENCE 572 AA; 67579 MW; A3DEA401 CRC32;  
 Query Match 71.0%; Score 49; DB 1; Length 572;  
 Best Local Similarity 85.7%; Pred. No. 4.08e+00;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Db 1 MPRGVAV 7  
 QY 1 MPRGVAV 7  
 RESULT 13  
 ID NAPA\_RHOSH STANDARD; PRT; 831 AA.  
 AC Q53176;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE PERIPLASMIC NITRATE REDUCTASE PRECURSOR (EC 1.7.99.4).  
 GN NAPA.  
 OS RHODOBACTER SPHAEROIDES (RHODOPSEUDOMONAS SPHAEROIDES).  
 OC PROKARYOTA; GRACILICUTES; ANOKYPHOTOBACTERIA; PURPLE BACTERIA;  
 OC RHODOSPIRILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DSM 158;  
 RX MEDLINE: 9633266.  
 RA REYES F., ROLDAN M.D., KLIPP W., CASTILLO F., MORENO-VIVIAN C.;  
 RL MOL. MICROBIOL. 19:1307-1318(1996).  
 CC -1- FUNCTION: LARGE SUBUNIT OF THE PERIPLASMIC NITRATE REDUCTASE  
 CC (NAP). ONLY EXPRESSED AT HIGH LEVELS DURING AEROBIC GROWTH. NAPA  
 CC COMPLEX RECEIVES ELECTRONS FROM THE MEMBRANE-ANCHORED TETRAHAEM  
 CC NACP PROTEIN, THUS ALLOWING ELECTRON FLOW BETWEEN MEMBRANE AND  
 CC PERIPLASM. ESSENTIAL FUNCTION FOR NITRATE ASSIMILATION AN MAY HAVE  
 CC A ROLE IN ANAEROBIC METABOLISM.  
 CC -1- CATALYTIC ACTIVITY: NITRITE + ACCEPTOR = NITRATE + REDUCED  
 CC ACCEPTOR.  
 CC -1- COFACTOR: BINDS THE GMP DERIVATIVE OF THE MOLYBDOPTERIN (MGD). MAY  
 CC BIND A 4FE-4S CLUSTER (BY SIMILARITY).  
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC SUBUNIT AND A CYTOCHROME C (BY  
 CC SIMILARITY).

```

CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -1- INDUCTION: NITRATE REDUCTASE ACTIVITY CAN BE INDUCED BY NITRATE
CC AND NOT REPRRESSD BY AMMONIUM OR OXYGEN.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC MOLYBDOPTERIN-CONTAINING
CC OXIDOREDUCTASES. BELONGS TO THE NASA/NAP/ANAB SUBFAMILY.
DR EMBL: Z46806; E186108;
DR PROSITE: PS00551; MOLYBDOPTERIN_PROK_1; 1.
DR PROSITE: PS00490; MOLYBDOPTERIN_PROK_2; FALSE NEG.
DR PROSITE: PS00932; MOLYBDOPTERIN_PROK_3; FALSE NEG.
KW NITRATE ASSIMILATION; OXIDOREDUCTASE; ELECTRON TRANSPORT; MOLYBDENUM;
KW PERIPLASMIC; SIGNAL; IRON-SULFUR; 4FE-4S.
FT SIGNAL 1 29
FT CHAIN 30 831 PERIPLASMIC NITRATE REDUCTASE.
FT DOMAIN 11 21 POLY-ALA.
FT METAL 48 48 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 51 51 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 55 55 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 83 83 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT NP_BIND 586 593 ATP (POTENTIAL).
SQ SEQUENCE 831 AA; 92806 MW; 740D8661 CRC32;

Query Match 71.0%; Score 49; DB 1; Length 831;
Best Local Similarity 100.0%; Pred. No. 4.08e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 788 MPRGV 793
OY 1 MPRGV 6
|||||
RESULT 14 STANDARD: PRT: 831 AA.
ID NAPA_PARDT
AC 056350;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PERIPLASMIC NITRATE REDUCTASE PRECURSOR (EC 1.7.99.4).
EN NAPA.
OS PARACOCCTUS DENITRIFICANS (SUBSP. THIOSPHAERA PANTOTROPHA).
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC UNCERTAIN.
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN-ATCC 35512 / LMD 82.5;
RX MEDLINE: 95366980.
RA BERKS B.C., RICHARDSON D.J., REILLY A., WILLIS A.C., FERGUSON S.J.;
RL BIOCHEM. J. 309:983-992(1995).
RN [2]
RP CHARACTERIZATION.
RC STRAIN-M-6;
RX MEDLINE: 94164150.
RA BERKS B.C., RICHARDSON D.J., ROBINSON C., REILLY A., APLIN R.T.,
RL EUR. J. BIOCHEM. 220:117-124(1994).
RN [3]
RP CHARACTERIZATION, AND SEQUENCE OF 642-655 AND 699-715.
RC STRAIN-M-6;
RX MEDLINE: 94252409.
RA BRETON J., BERKS B.C., REILLY A., THOMSON A.J., FERGUSON S.J.,
RL RICHARDSON D.J.;
FEBS LETT. 345:76-80(1994).
CC -1- FUNCTION: LARGE SUBUNIT OF THE PERIPLASMIC NITRATE REDUCTASE
CC (NAP): ONLY EXPRESSED AT HIGH LEVELS DURING AEROBIC GROWTH. NAPAB
CC COMPLEX RECEIVES ELECTRONS FROM THE MEMBRANE-ANCHORED TETRAHEM
CC NAPC PROTEIN, THUS ALLOWING ELECTRON FLOW BETWEEN MEMBRANE AND
CC PERIPLASM. ESSENTIAL FUNCTION FOR NITRATE ASSIMILATION AN MAY HAVE
CC A ROLE IN ANAEROBIC METABOLISM.
CC -1- CATALYTIC ACTIVITY: NITRITE + ACCEPTOR = NITRATE + REDUCED
CC ACCEPTOR.
CC -1- COFACTOR: BINDS THE GMP DERIVATIVE OF THE MOLYBDOPTERIN (MGD). MAY
CC BIND A 4FE-4S CLUSTER.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC SUBUNIT AND A CYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.

```

```

CC -1- SIMILARITY: TO OTHER PROKARYOTIC MOLYBDOPTERIN-CONTAINING
CC OXIDOREDUCTASES. BELONGS TO THE NASA/NAP/ANAB SUBFAMILY.
DR EMBL: Z36773; G600093;
DR PROSITE: PS00551; MOLYBDOPTERIN_PROK_1; 1.
DR PROSITE: PS00490; MOLYBDOPTERIN_PROK_2; FALSE NEG.
DR PROSITE: PS00932; MOLYBDOPTERIN_PROK_3; FALSE NEG.
KW NITRATE ASSIMILATION; OXIDOREDUCTASE; ELECTRON TRANSPORT; MOLYBDENUM;
KW PERIPLASMIC; SIGNAL; IRON-SULFUR; 4FE-4S.
FT SIGNAL 1 31
FT CHAIN 32 831 PERIPLASMIC NITRATE REDUCTASE.
FT METAL 48 48 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 51 51 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 55 55 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 83 83 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT CONFLICT 128 128 T -> D (IN AA SEQUENCE).
SQ SEQUENCE 831 AA; 92617 MW; DA6D314B CRC32;

Query Match 71.0%; Score 49; DB 1; Length 831;
Best Local Similarity 100.0%; Pred. No. 4.08e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 788 MPRGV 793
OY 1 MPRGV 6
|||||
RESULT 15 STANDARD: PRT: 54 AA.
ID YH1_STRCO
AC P16248;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 5.6 KD PROTEIN IN HISTIDINE BIOSYNTHESIS OPERON.
OS STREPTOMYCES COELICOLOR.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE: 90337345.
RA LIMAURO D., AVIRABILE A., CAPPELLANO M., PUGLIA A.M., BRUNI C.B.;
RL GENE 90:31-41(1990).
DR EMBL: M31628; G153299;
DR PIR: J00639; J00639.
KW HISTIDINE BIOSYNTHESIS; HYPOTHETICAL PROTEIN.
SQ SEQUENCE 54 AA; 5686 MW; 214B8704 CRC32;

Query Match 69.6%; Score 48; DB 1; Length 54;
Best Local Similarity 55.6%; Pred. No. 6.74e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 27 MPRGLIYVL 35
OY 1 MPRGVVTL 9
|||||

```

Search completed: Fri Sep 11 13:22:20 1998  
 Job time : 8 secs.



(TM)

Distribution rights by Oxford Molecular Ltd

```
Run on:      Fri Sep 11 13:22:38 1998; MasPar time 3.72 Seconds
```

1 MPRGVVITL 9

Gap 15

140555 seqs, 42109429 residues

Listing first 45 summaries

13:sp\_unclassified

Mean 23.437; Variance 24.835; scale 0.944

and is derived by analysis of the total score distribution.

## SUMMARIES

| No. | Score | Query Match | Length | DB | ID     | Description            | Pred. No. |
|-----|-------|-------------|--------|----|--------|------------------------|-----------|
| 1   | 53    | 76.6        | 714    | 9  | 007170 | FUSAZ.                 | 4.66e-01  |
| 2   | 42    | 71.0        | 314    | 8  | P93459 | 1-AMINOCYCLOPROPANE-1- | 4.11e+00  |
| 3   | 49    | 71.0        | 470    | 8  | 001487 | COSMID C13F10.         | 4.11e+00  |
| 4   | 49    | 71.0        | 697    | 11 | 087026 | CAPSID.                | 4.11e+00  |
| 5   | 49    | 71.0        | 1154   | 1  | 001375 | CONTAINS REVERSE TRANS | 4.11e+00  |
| 6   | 49    | 71.0        | 1154   | 1  | 001379 | CONTAINS REVERSE TRANS | 4.11e+00  |
| 7   | 49    | 71.0        | 2658   | 8  | 030914 | DAPWOMYCIN BIOSYNTHETI | 4.11e+00  |
| 8   | 48    | 69.6        | 205    | 8  | 041860 | TRANSPOSABLE ELEMENT M | 6.94e+00  |
| 9   | 48    | 69.6        | 346    | 9  | 025017 | GERC2 PROTEIN (GERC2). | 6.94e+00  |
| 10  | 48    | 69.6        | 338    | 10 | 035400 | HYDROXYSTEROID SULFOIR | 6.94e+00  |
| 11  | 48    | 69.6        | 350    | 2  | 000204 | HYDROXYSTEROID SULFOIR | 6.94e+00  |
| 12  | 48    | 69.6        | 365    | 2  | 000205 | HYDROXYSTEROID SULFOIR | 6.94e+00  |
| 13  | 48    | 69.6        | 526    | 9  | 0Q5679 | SIMILARITY TO GALACTOS | 6.94e+00  |
| 14  | 48    | 69.6        | 538    | 4  | 028939 | VASCULAR CELL ADHESION | 6.94e+00  |
| 15  | 48    | 69.6        | 538    | 4  | 029123 | VASCULAR CELL ADHESION | 6.94e+00  |
| 16  | 48    | 69.6        | 550    | 3  | 011145 | GAMMA-AMINOBUTYRIC ACT | 6.94e+00  |
| 17  | 48    | 69.6        | 562    | 9  | P71605 | HYPOTHETICAL 61.0 KD P | 6.94e+00  |
| 18  | 47    | 68.1        | 151    | 11 | 040633 | INTERLEUKIN 17.        | 1.16e+01  |
| 19  | 47    | 68.1        | 271    | 1  | 034969 | MITOGENIC FACTOR PRECU | 1.16e+01  |
| 20  | 47    | 68.1        | 339    | 3  | 011007 | CATHEPSIN B-LIKE CYSTE | 1.16e+01  |

|    |    |      |      |         |                        |          |
|----|----|------|------|---------|------------------------|----------|
| 45 | 45 | 65.2 | 1520 | 0.15829 | CARBAMYL PHOSPHATE SYN | 1.16e+01 |
| 44 | 45 | 65.2 | 1083 | 0.15829 | CARBAMYL PHOSPHATE SYN | 1.16e+01 |
| 43 | 45 | 65.2 | 1068 | 0.15829 | CARBAMYL PHOSPHATE SYN | 1.16e+01 |
| 42 | 45 | 65.2 | 1068 | 0.15829 | CARBAMYL PHOSPHATE SYN | 1.16e+01 |
| 41 | 45 | 65.2 | 956  | 0.16099 | EXCITATORY AMINO ACID  | 3.18e+01 |
| 40 | 45 | 65.2 | 542  | 0.30147 | LONG-CHAIN-FATTY-ACID- | 3.18e+01 |
| 39 | 45 | 65.2 | 501  | 0.26978 | ATP-DEPENDENT PROTEASE | 3.18e+01 |
| 38 | 45 | 65.2 | 476  | 0.26978 | ATP-DEPENDENT PROTEASE | 3.18e+01 |
| 37 | 45 | 65.2 | 202  | 0.06273 | GGAH.                  | 3.18e+01 |
| 36 | 45 | 65.2 | 79   | 0.11361 | SMILAR TO VARIOA A11   | 3.18e+01 |
| 35 | 46 | 66.7 | 1067 | 0.03867 | P123.                  | 1.93e+01 |
| 34 | 46 | 66.7 | 651  | 0.02070 | SMILARITY TO COA METH  | 1.93e+01 |
| 33 | 46 | 66.7 | 382  | 0.21998 | GALACTOKINASE.         | 1.93e+01 |
| 32 | 46 | 66.7 | 341  | 0.05475 | HYPOHETICAL 37.5 KD P  | 1.93e+01 |
| 31 | 46 | 66.7 | 243  | 0.17308 | GBA RECEPTOR SUBUNIT   | 1.93e+01 |
| 30 | 46 | 66.7 | 215  | 0.14324 | HYPOHETICAL 24.8 KD P  | 1.93e+01 |
| 29 | 46 | 66.7 | 187  | 0.03295 | RIBOSOMAL PROTEIN U5.  | 1.93e+01 |
| 28 | 46 | 66.7 | 99   | 0.02562 | HYPOHETICAL 11.0 KD P  | 1.93e+01 |
| 27 | 46 | 66.7 | 64   | 0.01875 | ORL1 (FRAGMENT)        | 1.93e+01 |
| 26 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 25 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 24 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 23 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 22 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 21 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 20 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 19 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 18 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 17 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 16 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 15 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 14 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 13 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 12 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 11 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 10 | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 9  | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 8  | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 7  | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 6  | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 5  | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 4  | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 3  | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 2  | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |
| 1  | 47 | 68.1 | 2183 | 0.09833 | RNA DEPENDENT RNA POLY | 1.16e+01 |

## ALIGNMENTS

|                       |  |              |       |               |
|-----------------------|--|--------------|-------|---------------|
| RESULT                | 1  | PRELIMINARY; | PRT;  | 714 AA.       |
| ID                    | 007170   |              |       |               |
| AC                    | 007170;  |              |       |               |
| DT                    | 01-JUL-1997 (TREMBLREL. 04, CREATED)                               |              |       |               |
| DT                    | 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)                  |              |       |               |
| DT                    | 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)                |              |       |               |
| DE                    | FUSA2.   |              |       |               |
| GN                    | FUSA2.   |              |       |               |
| OS                    | MYCOBACTERIUM TUBERCULOSIS.  |              |       |               |
| OC                    | PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.         |              |       |               |
| RN                    | [1]  |              |       |               |
| RP                    | SEQUENCE FROM N.A.   |              |       |               |
| RC                    | STRAIN-H37RV;  |              |       |               |
| RA                    | OLIVER K., HARRIS D.;  |              |       |               |
| RL                    | SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.               |              |       |               |
| RN                    | [2]  |              |       |               |
| RP                    | SEQUENCE FROM N.A.   |              |       |               |
| RC                    | STRAIN-H37RV;  |              |       |               |
| RA                    | PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;                        |              |       |               |
| RL                    | SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.               |              |       |               |
| RN                    | [3]  |              |       |               |
| RP                    | SEQUENCE FROM N.A.   |              |       |               |
| RC                    | STRAIN-H37RV;  |              |       |               |
| RA                    | MEDLINE: 96181548.   |              |       |               |
| RX                    | PHILIPP W.J., POULET S., EIGMEIER K., PASCOPELLA L.,               |              |       |               |
| RA                    | BRASLOBRMANIAN V., HEYM B., BENGH S., BLOOM B.R., JACOBS W.R. JR., |              |       |               |
| RA                    | COLE S.T.;   |              |       |               |
| RL                    | PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).                  |              |       |               |
| DR                    | EMBL: 296071; E321091; -   |              |       |               |
| SO                    | SEQUENCE 714 AA; 75630 MW; D3E5E4E8 CRC32;                         |              |       |               |
| Query Match           |  |              |       |               |
| Best Local Similarity | 76.8%;   | Score 53;    | DB 9; | Length 714;   |
| Matches               | 6;   | Conservative | 2;    | Mismatches 0; |
| Indels                | 0;   | Gaps         | 0;    |               |
| Db                    | 141 MPRAVIT 148  |              |       |               |
| OY                    | 1 MPRGVYT 8  |              |       |               |
| RESULT                | 2  | PRELIMINARY; | PRT;  | 314 AA.       |
| ID                    | P93459   |              |       |               |
| AC                    | P93459;  |              |       |               |
| DT                    | 01-MAY-1997 (TREMBLREL. 03, CREATED)                               |              |       |               |

DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)  
DE 1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID OXIDASE.  
OS RUMEX PALUSTRIS.  
OC EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; VIRIDIPALANTAE;  
OC CHAROPHYTA/EMBRIOPHYTA GROUP; EMBRIOPHYTA; MAGNOLIOPHYTA;  
OC MAGNOLIOPSIDA; THEANAEE; POLYGONALES; POLYGONACEAE; RUMEX.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LEAF;  
RA VRIEZEN H.W., HULZINK R.;  
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Y10034; E286173;  
SQ SEQUENCE 314 AA; 35665 MW; 5E318BA7 CRC32;

Query Match 71.0%; Score 49; DB 8; Length 314;  
Best Local Similarity 55.6%; Pred. No. 4.11e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 208 MPRSVVNL 216  
|:|:|:|:|

Oy 1 MPRGVVTL 9

RESULT 3  
ID 001487 PRELIMINARY; PRT; 470 AA.  
AC 001487;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DE 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
DE COSMID C13F10.  
GN C13F10.6.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE: 94150718;  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKES M.,  
RA BONFIELD J., BURTON M., GONNELL M., CORSEY T., COOPER J.,  
RA COULSON A., CRAXTON M., DEAR S., DU Z., DUBBIN R., FAVELLO A.,  
RA FULTON L., GARDNER A., GREEN P., HAMKINS T., HILLIER L., JIER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,  
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFEEN L., ROOPRA A.,  
RA SAUNDERS D., SHOMKEEN R., SMALDON N., SMITH A., SONNHAMMER E.,  
RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA TTN A., WOHLDMANN P.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U97006; G1938463;  
SQ SEQUENCE 470 AA; 54820 MW; 3D5D34F CRC32;

Query Match 71.0%; Score 49; DB 3; Length 470;  
Best Local Similarity 44.4%; Pred. No. 4.11e+00;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 228 MPRNLIAL 236  
|:|:|:|:|

Oy 1 MPRGVVTL 9

RESULT 4

ID 087026 PRELIMINARY; PRT; 697 AA.  
AC 087026;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE CAPSID.  
GN CAP.  
OS SACCHAROMYCES CEREVISIAE VIRUS LA.  
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94111988.  
RA BRUENN J.A.;  
RL NUCLEIC ACIDS RES. 21:5667-5669(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA PARK C.M., LOPINSKI J.D., MASUDA J., TZENG T.H., BRUENN J.A.;  
RL VIROLOGY 216:451-454(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA BRUENN J.A.;  
RL SUBMITTED (AUG-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U01060; G595250;  
SQ SEQUENCE 697 AA; 78315 MW; 5C1BD5E2 CRC32;

Query Match 71.0%; Score 49; DB 11; Length 697;  
Best Local Similarity 55.6%; Pred. No. 4.11e+00;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 393 MNRGITVL 401  
|:|:|:|:|

Oy 1 MPRGVVTL 9

RESULT 5  
ID 001375 PRELIMINARY; PRT; 1154 AA.  
AC 001375;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE CONTRAINS REVERSE TRANSCRIPTASE AND CYS FINGER DOMAINS.  
OS NEUROSPORA CRASSA.  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; PYRENOYCETES; SORDARIALES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-01518;  
RX MEDLINE: 94203179.  
RA CAMBARERI E.B., HELBER J., KINSEY J.A.;  
RL MOL. GEN. GENET. 242:658-665(1994).  
DR EMBL: L25662; G409761;  
KW RNA-DIRECTED DNA POLYMERASE.  
SQ SEQUENCE 1154 AA; 130399 MW; DF0BA680 CRC32;

Query Match 71.0%; Score 49; DB 1; Length 1154;  
Best Local Similarity 55.6%; Pred. No. 4.11e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 828 KPRGIVTL 836  
|:|:|:|:|

Oy 1 MPRGVVTL 9

RESULT 6  
ID 001379 PRELIMINARY; PRT; 1154 AA.  
AC 001379;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE CONTRAINS REVERSE TRANSCRIPTASE AND CYS FINGER DOMAINS.  
OS NEUROSPORA CRASSA.  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; PYRENOYCETES; SORDARIALES.  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-J1518;  
 RX MEDLINE: 94203179.  
 RA CAMBARERI E.B., HELBER J., KINSEY J.A.;  
 RL MOL. GEN. GENET. 242:658-665(1994).  
 DR EMBL: L25663; G409764.  
 KW RNA-DIRECTED DNA POLYMERASE.  
 SQ SEQUENCE 1154 AA; 130471 MW; 7EBE8EAF CRC32;

Query Match 71.0%; Score 49; DB 1; Length 1154;  
 Best Local Similarity 55.6%; Pred. No. 4.11e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 828 KPRGIVGL 836  
 QY 1 MPRGVVTL 9

RESULT 7  
 ID 030914; PRELIMINARY; PRT; 2638 AA.  
 AC 030914;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DE DAPTOMYCIN BIOSYNTHETIC PROTEIN SUBUNIT (FRAGMENT).  
 OS STREPTOMYCES ROSEOSPORUS.  
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.  
 RN [1]  
 RC STRAIN-A21978.6, A21978.65;  
 RA MCHENNEY M.A., HOSTED T.J., DEHOFF B.S., ROSTECK P.R. JR., BALTZ R.H.;  
 RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AF021262; G2501965.  
 DE PROSITE: PS00455; AMP-BINDING; 2.  
 FT NON\_TER 1 1  
 FT NON\_TER 2638 2638  
 SQ SEQUENCE 2638 AA; 282397 MW; 922B838D CRC32;

Query Match 71.0%; Score 49; DB 9; Length 2638;  
 Best Local Similarity 75.0%; Pred. No. 4.11e+00;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 2410 LPRGVVTL 2417  
 QY 1 MPRGVVTL 8

RESULT 8  
 ID 041860; PRELIMINARY; PRT; 206 AA.  
 AC 041860;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE TRANSPOSABLE ELEMENT MUI SEQUENCE.  
 OS ZEA MAYS (MAIZE).  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;  
 OC CYPERALES; GRAMINEAE.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RX TRANSPOSON-ZEA MAYS;  
 RX MEDLINE: 84297205.  
 RA BARKER R.F., THOMPSON D.V., TALBOT D.R., SWANSON J.;  
 RA BENNETZEN J.L.;  
 RL NUCLEIC ACIDS RES. 12:5955-5967(1984).  
 DR EMBL: X00913; G22496; -;  
 SQ SEQUENCE 206 AA; 21528 MW; 4356B562 CRC32;

Query Match 69.6%; Score 48; DB 8; Length 206;  
 Best Local Similarity 66.7%; Pred. No. 6.94e+00;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 149 VPRGVVTL 157  
 QY 1 MPRGVVTL 9

RESULT 9  
 ID 026017; PRELIMINARY; PRT; 246 AA.  
 AC 026017;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE GERC2 PROTEIN (GERC2).  
 GN HP1483.  
 OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;  
 OC AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-26695;  
 RA TOMB, WHITE, KERLAVAGE, CLAYTON, SUTTON, FLEISCHMANN, KETCHUM, KLENK,  
 RA GILL, DOUGHERTY, NELSON, QUACKENBUSH, ZHOU, KIRKNESS, PETERSON, LOFTUS,  
 RA RICHARDSON, DODSON, KHALAK, GLODER, MCKENNEY, FITZGERALD, LEE, ADAMS,  
 RA HICKEY, BEGG, GOCAYNE, UTTERBACK, PETERSON, KELLEY, COTTON, WEIDMAN,  
 RA FUJII, BOWMAN, WATTHEY, WALLIN, HAYES, BORODOVSKY, KARP, SMITH,  
 RA FRASER VENTER;  
 RL NATURE 388:539-547(1997).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-26695;  
 RA TOMB, WHITE, KERLAVAGE, CLAYTON, SUTTON, FLEISCHMANN, KETCHUM, KLENK,  
 RA GILL, DOUGHERTY, NELSON, QUACKENBUSH, ZHOU, KIRKNESS, PETERSON, LOFTUS,  
 RA RICHARDSON, DODSON, KHALAK, GLODER, MCKENNEY, FITZGERALD, LEE, ADAMS,  
 RA HICKEY, BEGG, GOCAYNE, UTTERBACK, PETERSON, KELLEY, COTTON, WEIDMAN,  
 RA FUJII, BOWMAN, WATTHEY, WALLIN, HAYES, BORODOVSKY, KARP, SMITH,  
 RA FRASER VENTER;  
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AE000647; G2314655;  
 SQ SEQUENCE 246 AA; 27871 MW; 4DA2675D CRC32;

Query Match 69.6%; Score 48; DB 9; Length 246;  
 Best Local Similarity 66.7%; Pred. No. 6.94e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 155 KPRGVVTL 163  
 QY 1 MPRGVVTL 9

RESULT 10  
 ID 035400; PRELIMINARY; PRT; 338 AA.  
 AC 035400;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE HYDROXYSTEROID SULFOTRANSFERASE.  
 GN SUL2B.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RA SAKAKIBARA Y., YANAGISAWA K., TAKAMI Y., NAKAYAMA T., SUIKO M.,  
 RA LIU M.-C.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AF026072; G2570898; -;  
 KW TRANSFERASE.  
 SQ SEQUENCE 338 AA; 38407 MW; DD62738C CRC32;

Query Match 69.6%; Score 48; DB 10; Length 338;  
 Best Local Similarity 75.0%; Pred. No. 6.94e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 146 PRGVVSL 153  
 QY 2 PRGVVTL 9

```

RESULT 11
ID 000204 PRELIMINARY: PRT: 350 AA.
AC 000204:
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DE 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)
DE HYDROXYSTEROID SULFOTRANSFERASE HST12A.
GN HST12
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA;
RA HER C., WOOD T.C., EICHLER E., MOHREWEISER H.W., SICILIANO M.J.,
RA RAFTOGIANIS R.B., WEINSHILBOUM R.M.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U92314; G1923291; -
KW TRANSFERASE.
SQ SEQUENCE 350 AA: 39626 MW: 180E8E65 CRC32:

Query Match 69.6%; Score 48; DB 2; Length 350;
Best Local Similarity 75.0%; Pred. No. 6.94e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 134 PRDYYVSL 141
QY 2 PRGVVTL 9

RESULT 12
ID 000205 PRELIMINARY: PRT: 365 AA.
AC 000205:
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DE 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)
DE HYDROXYSTEROID SULFOTRANSFERASE HST12B.
GN HST12
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA;
RA HER C., WOOD T.C., EICHLER E., MOHREWEISER H.W., SICILIANO M.J.,
RA RAFTOGIANIS R.B., WEINSHILBOUM R.M.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U92315; G1923293; -
KW TRANSFERASE.
SQ SEQUENCE 365 AA: 41307 MW: 2E641641 CRC32:

Query Match 69.6%; Score 48; DB 2; Length 365;
Best Local Similarity 75.0%; Pred. No. 6.94e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 149 PRDYYVSL 156
QY 2 PRGVVTL 9

RESULT 13
ID 053679 PRELIMINARY: PRT: 526 AA.
AC 053679:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE SIMILARITY TO GALACTOSE OXIDASE FROM DACTYLITIUM DENDROIDES.
GN FEBB GENE.
OS STIGMATELLA AURANTIACA.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; MYXOBACTERIALES;
OC CYSTOBACTERACEAE.
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-DW4/3-1;
RA SILAKOWSKI B., POSPIECH A., NEUMANN B., SCHAIRER H.U.;
RL J. BACTERIOL. 178:6706-6713(1996).
DR EMBL: 211601; E245931; -
SQ SEQUENCE 526 AA: 57820 MW: 42FA610C CRC32:

Query Match 69.6%; Score 48; DB 9; Length 526;
Best Local Similarity 66.7%; Pred. No. 6.94e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 4 LPRGVVSL 12
QY 1 MPRGVVTL 9

RESULT 14
ID 028939 PRELIMINARY: PRT: 538 AA.
AC 028939:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE VASCULAR CELL ADHESION MOLECULE PRECURSOR.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; ARTIODACTYLA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-AORTA;
RX MEDLINE: 94271236.
RA TSANG Y.T., HASKARD D.O., ROBINSON M.K.;
RL BIOCHEM. BIOPHYS. RES. COMMUN. 201:805-812(1994).
DR EMBL: U08351; G474383; -
KW SIGNAL.
FT CHAIN 25 24 POTENTIAL.
SQ SEQUENCE 538 AA: 58795 MW: 6E5EA776 CRC32:

Query Match 69.6%; Score 48; DB 4; Length 538;
Best Local Similarity 55.6%; Pred. No. 6.94e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1 MPRNTVIE 9
QY 1 MPRGVVTL 9

RESULT 15
ID 029123 PRELIMINARY: PRT: 538 AA.
AC 029123:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE VASCULAR CELL ADHESION MOLECULE.
GN VCAM.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; ARTIODACTYLA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TNF-STIMULATED AORTIC;
RX MEDLINE: 94271236.
RA TSANG Y.T., HASKARD D.O., ROBINSON M.K.;
RL BIOCHEM. BIOPHYS. RES. COMMUN. 201:805-812(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-TNF-STIMULATED AORTIC;
RX MEDLINE: 96106451.
RA MOELLER J.P., EVANS M.J., COFFELL R., ROTHER R.P., MATIS L.A.,
RA ELIOTT E.A.;
RL TRANSPLANTATION 60:1299-1306(1995).
DR EMBL: L43124; G1199461; -
SQ SEQUENCE 538 AA: 58713 MW: 9401CC8A CRC32:

```

Sun Sep 13 10:55:20 1998

US-08-452-843-12.rspt

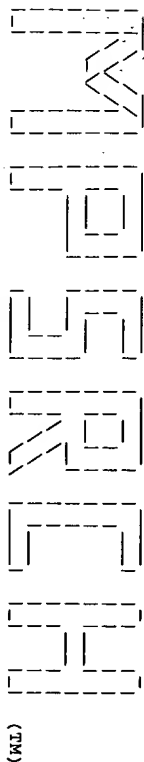
Page 5

Query Match: 69.68; Score 48; DB 4; Length 538;  
Best Local Similarity 55.68; Pred. NO. 6.94e+00;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Dp: 1 MPRNIVIF 9  
||: ||:  
QY: 1 MPRGVVTL 9

Search completed: Fri Sep 11 13:23:15 1998  
Time: 87 secs

**THIS PAGE BLANK (USPTO)**



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPearch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:17:24 1998; Maspar time 2.56 Seconds  
56.808 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-11  
Description: (1-9) from US08452843.pep  
Perfect Score: 69  
Sequence: 1 RYRGTVAL 9

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

a-geneseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 17.208; Variance 43.975; scale 0.391

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description           | Pred. No. |
|------------|-------|-------------|--------|----|--------|-----------------------|-----------|
| 1          | 69    | 100.0       | 9      | 18 | R89372 | Histone H3.3 derived  | 1.14e+01  |
| 2          | 69    | 100.0       | 70     | 10 | R46075 | Histone H3.3 like pro | 1.14e+01  |
| 3          | 47    | 68.1        | 19     | 14 | R74668 | H1A-DB1*0405 binding  | 7.76e+01  |
| 4          | 47    | 68.1        | 19     | 15 | R83693 | H1A binding peptide h | 7.76e+01  |
| 5          | 46    | 66.7        | 16     | 1  | R03457 | Intracellular retenti | 1.02e+02  |
| 6          | 46    | 66.7        | 332    | 20 | W06491 | Beta-1-4-galactosyltr | 1.02e+02  |
| 7          | 46    | 66.7        | 441    | 1  | R05272 | Polypeptide with amin | 1.02e+02  |
| 8          | 46    | 66.7        | 472    | 20 | R97245 | Virulence gene cluste | 1.02e+02  |
| 9          | 45    | 65.2        | 31     | 4  | R21422 | Matrix peptide from b | 1.35e+02  |
| 10         | 45    | 65.2        | 52     | 3  | R12876 | Non-collagenous bone  | 1.35e+02  |
| 11         | 45    | 65.2        | 52     | 3  | R12875 | Non-collagenous bone  | 1.35e+02  |
| 12         | 45    | 65.2        | 239    | 16 | R89423 | Mucin-derived protein | 1.35e+02  |
| 13         | 45    | 65.2        | 240    | 16 | R89422 | Mucin-derived protein | 1.35e+02  |
| 14         | 45    | 65.2        | 255    | 16 | R89420 | Mucin-derived protein | 1.35e+02  |
| 15         | 45    | 65.2        | 254    | 16 | R89421 | Mucin-derived protein | 1.35e+02  |
| 16         | 45    | 65.2        | 273    | 16 | R89418 | Mucin-derived protein | 1.35e+02  |
| 17         | 45    | 65.2        | 282    | 16 | R89419 | Mucin-derived protein | 1.35e+02  |
| 18         | 45    | 65.2        | 327    | 17 | R96298 | Glycoprotein 39 C ter | 1.35e+02  |

|    |    |      |      |    |        |                        |          |
|----|----|------|------|----|--------|------------------------|----------|
| 19 | 45 | 65.2 | 348  | 4  | R27662 | C-terminal region of   | 1.35e+02 |
| 20 | 45 | 65.2 | 455  | 4  | R23973 | Transmembrane form of  | 1.35e+02 |
| 21 | 45 | 65.2 | 497  | 18 | R93184 | Human cytochrome P450  | 1.35e+02 |
| 22 | 45 | 65.2 | 497  | 18 | R72378 | Human auxiliary cytoch | 1.35e+02 |
| 23 | 45 | 65.2 | 497  | 18 | R93183 | Human auxiliary cytoch | 1.35e+02 |
| 24 | 45 | 65.2 | 497  | 18 | R93182 | Human cytochrome P450  | 1.35e+02 |
| 25 | 45 | 65.2 | 497  | 17 | R81462 | Human cytochrome P450  | 1.35e+02 |
| 26 | 45 | 65.2 | 497  | 13 | R72376 | Human derived cytochr  | 1.35e+02 |
| 27 | 45 | 65.2 | 497  | 13 | R72375 | Human auxiliary cytoch | 1.35e+02 |
| 28 | 45 | 65.2 | 497  | 13 | R72377 | Human auxiliary cytoch | 1.35e+02 |
| 29 | 45 | 65.2 | 497  | 13 | R93185 | Human auxiliary cytoch | 1.35e+02 |
| 30 | 45 | 65.2 | 523  | 12 | R71976 | Human cytochrome P450  | 1.35e+02 |
| 31 | 45 | 65.2 | 770  | 28 | W34199 | Streptomyces efflux p  | 1.35e+02 |
| 32 | 45 | 65.2 | 824  | 24 | W23774 | Bordetella pertussis   | 1.35e+02 |
| 33 | 45 | 65.2 | 1528 | 27 | W33363 | Human multidrug resis  | 1.35e+02 |
| 34 | 45 | 65.2 | 3164 | 16 | R94345 | Hepatitis GB virus (H  | 1.76e+02 |
| 35 | 44 | 63.8 | 655  | 6  | R31041 | Simr polypeptide.      | 1.76e+02 |
| 36 | 44 | 63.8 | 678  | 8  | R42087 | D. melanogaster dorsa  | 1.76e+02 |
| 37 | 44 | 63.8 | 2466 | 13 | R71498 | Human protein tyrosin  | 1.76e+02 |
| 38 | 43 | 62.3 | 246  | 12 | R62753 | Seib sequence.         | 2.31e+02 |
| 39 | 43 | 62.3 | 322  | 16 | R48754 | Rat RGH G-protein cou  | 2.31e+02 |
| 40 | 43 | 62.3 | 322  | 19 | W02726 | Rat RGH G-protein c    | 2.31e+02 |
| 41 | 43 | 62.3 | 391  | 7  | R39259 | Human somatostatin re  | 2.31e+02 |
| 42 | 43 | 62.3 | 391  | 7  | R39260 | Murine somatostatin r  | 2.31e+02 |
| 43 | 43 | 62.3 | 393  | 19 | W04245 | Human G-protein coupl  | 2.31e+02 |
| 44 | 43 | 62.3 | 489  | 23 | W01556 | Acetabone C-11 hydro   | 2.31e+02 |
| 45 | 43 | 62.3 | 826  | 5  | R26042 | P. yoellii SSP2 antige | 2.31e+02 |

## ALIGNMENTS

RESULT 1  
ID R89372 standard; peptide: 9 AA.

AC R89372:  
DE Histone H3.3 derived immunogenic peptide.  
KW Immunogenic peptide; supermotif: H1A molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN W09603140-A1.  
PD 08-FEB-1996.  
PF 21-FEB-1995; U09234.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J.  
DR WPI: 96-116784/12.  
PT Compn. comprising immunogenic peptide with supermotif allowing more  
than one H1A mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PT Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
use in the composition of the invention. The composition comprises  
an immunogenic peptide of 9-10 residues with a supermotif which  
allows binding of more than one H1A molecule. It pref. comprises  
two conserved residues, a first at the 2nd position from the N-  
terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
are used to induce a CTL response in a patient. They are also  
useful in compositions for in vivo and ex vivo therapeutic and  
diagnostic applications, e.g. the treatment of cancer and viral  
infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 100.0%; Score 69; DB 18; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.14e+01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 RYRGTVAL 9  
|||  
OY 1 RYRGTVAL 9

RESULT 2  
 AC R46075: standard: Protein; 70 AA.  
 ID R46075:  
 DE 19-OCT-1994 (first entry)  
 DT Histone H3.3 like protein.  
 KW Human CDNA; library: enzyme; protein.  
 OS Homo sapiens.  
 PN M09403599-A.  
 PD 17-FEB-1994.  
 PF 04-AUG-1993; J01095.  
 PR 04-AUG-1992; JP-208077.  
 PR 13-NOV-1992; JP-327619.  
 PR 26-FEB-1993; JP-061431.  
 PA (SAGA) SAGAMI CHEM RES CENTRE.  
 PI Iwahori A, Kato S, Kato T, Kim N, Oh S, Sekine S;  
 DR WPI: 94-065688/08.  
 DR N-PSDB: 057414.  
 PT cDNA of human origin and proteins coded by it - which may be  
 PT expressed by in vivo or in vitro translation using sense RNA or  
 PT antisense DNA corresponding to the cDNA.  
 PS Claim 1, Page 29; 167pp. Japanese.  
 CC mRNA expressed in human fibrosarcoma cell line HT-1080 was  
 CC isolated and used to construct a cDNA library using vector  
 CC pXa1. Clone HP00014 encoding histone H3.3-like protein  
 CC was isolated.  
 SQ Sequence 70 AA;

Query Match 100.0%; Score 69; DB 10; Length 70;  
 Best Local Similarity 100.0%; Pred. No. 1.14e+01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 41 YRPCTVAL 49  
 |||||  
 OY 1 YRPCTVAL 9

RESULT 3  
 ID R74668 standard: peptide; 19 AA.  
 AC R74668:  
 DE 19-DEC-1995 (first entry)  
 DT HLA-DRB1\*0405 binding oligopeptide (VI).  
 KW Oligopeptide; HLA-DRB1\*0405; immunosuppressant; lymphocyte;  
 KW Epstein-Barr virus; B cell line.  
 OS Homo sapiens.  
 PN J07082295-A.  
 PD 28-MAR-1995.  
 PF 13-SEP-1993; 227091.  
 PR 13-SEP-1993; JP-227091.  
 PA (TEIJ) TEIJIN LTD.  
 DR WPI: 95-158991/21.  
 PT Oligopeptide immunosuppressant - isolated from B lymphocytes of  
 PT HLA-DRB1\*0405-subjects or prep. by peptide synthesis  
 PS Claim 2, Page 2; 8pp. Japanese.  
 CC The sequences given in R74663-68 represent oligopeptides which bind  
 CC to HLA-DRB1\*0405. These peptides act as immunosuppressants and are  
 CC administered at a daily dose of 1-100 mg/kg. These peptides may be  
 CC derived from lymphocytes derived from a patient having HLA-DRB1\*0405  
 CC and treated with Epstein-Barr virus to give a B cell line to produce  
 CC the peptides.  
 SQ Sequence 19 AA;

Query Match 68.1%; Score 47; DB 14; Length 19;  
 Best Local Similarity 75.0%; Pred. No. 7.76e+01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 6 YRPCTVAL 13  
 |||||  
 OY 2 YRPCTVAL 9

RESULT 4  
 ID R83693 standard: peptide; 19 AA.

AC R83693;  
 DT 10-APR-1996 (first entry)  
 DE HLA binding peptide homologous to pyruvate kinase M2 isozyme.  
 KW HLA binding oligopeptide; immunosuppressant; autoimmune disease;  
 KW pyruvate kinase; M2 isozyme; residues 101-119; homologue.  
 OS Synthetic.  
 PN J07206896-A.  
 PD 08-AUG-1995.  
 PF 20-JAN-1994; 004615.  
 PR 20-JAN-1994; JP-004615.  
 PA (TEIJ) TEIJIN LTD.  
 DR WPI: 95-309097/40.  
 PT New HLA binding oligo:peptide(s) - useful as immunosuppressants for  
 PT treating autoimmune diseases  
 PS Example 1, Page 5; 9pp. Japanese.  
 CC The present peptide is homologous to the pyruvate kinase M2 isozyme  
 CC residues 101-119, and is a HLA binding oligopeptide. It can be used  
 CC as an immunosuppressant for the treatment of autoimmune diseases.  
 SQ Sequence 19 AA;

Query Match 68.1%; Score 47; DB 15; Length 19;  
 Best Local Similarity 75.0%; Pred. No. 7.76e+01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 6 YRPCTVAL 13  
 |||||  
 OY 2 YRPCTVAL 9

RESULT 5  
 ID R03457 standard: protein; 16 AA.  
 AC R03457:  
 DT 02-AUG-1990 (first entry)  
 DE Intracellular retention moiety derived from Thymine Z-1.  
 KW Intracellular retention moiety; Thymine Z-1; tumour therapy.  
 OS Synthetic.  
 PN EP-359347-A.  
 PD 21-MAR-1990.  
 PF 14-AUG-1989; 250014.  
 PR 15-AUG-1988; US-232337.  
 PA (NEOR-) Neorx Corp.  
 PI Anderson DC, Morgan AC, Abrams PG, Nichols EJ, Fritzbeg AR;  
 DR WPI: 90-085154/12.  
 PT Covalently linked complex for tumour treatment - comprises  
 PT treating with protein, cytotoxic agent and enhancing moiety.  
 PS Claim 16; Page 22; 23pp. English.  
 CC The sequence is one of several possible intracellular retention moieties  
 CC which can be covalently attached to one or more other enhancing moieties  
 CC such as an internalization moiety, and to a targeting protein and a  
 CC cytotoxic agent. The moiety is designed to bind noncovalently to dDNA  
 CC in the cell so increasing the amt. of time that the targeting protein  
 CC conjugate is retained intracellularly. The N-terminal Cys and Gly  
 CC residues are added to allow covalent cross linking to the targeting  
 CC protein. The C-terminal is amidated. The complex is useful for treatment  
 CC and diagnosis of tumours.  
 CC See also R03435-60.  
 SQ Sequence 16 AA;

Query Match 66.7%; Score 46; DB 1; Length 16;  
 Best Local Similarity 75.0%; Pred. No. 1.02e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 9 YRPCTVAL 16  
 |||||  
 OY 1 YRPCTVAL 8

RESULT 6  
 ID W06491 standard: Protein; 332 AA.  
 AC W06491;  
 DT 05-FEB-1997 (first entry)  
 DE Beta-1-4-galactosyltransferase-related protein #2.  
 KW Murine; beta-1-4-galactosyltransferase-related protein; sterility;



KM fertilisation; F9 cancer cell; Huynh's method.  
 OS Mus musculus.  
 PD 06-AUG-1996.  
 PF 25-JAN-1995: 009642.  
 PR 25-JAN-1995: JP-009642.  
 PA (MITK) MITSUI TOAISO CHEM INC.  
 PA (MORA/) MORAMATSU T.  
 DR N-PSDB: T45082.  
 PT DNA sequence encoding beta-1,4-galactosyl:transferase-related protein - useful for sterility diagnosis; and for assisting or inhibiting fertilisation.  
 PS Claim 4; Page 7-9; 11pp: Japanese.  
 CC The sequences given in W06490-91 represent two clones of murine beta-1,4-galactosyltransferase-related proteins. These proteins can be used as diagnostic agents for various diseases. They are esp. useful in the diagnosis of sterility and in the aiding and inhibiting of fertilisation. The cDNA's encoding the two beta-1,4-galactosyltransferase-related proteins were isolated from F9 cancer cells according to Huynh's method.  
 SO Sequence 332 AA.  
 Query Match  
 Best Local Similarity 66.7%; Score 46; DB 20; Length 332;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 39 frpdyvdl 46  
 |||||  
 OY 2 YRPTVAL 9  
 RESULT  
 ID R05272 standard; protein; 441 AA.  
 AC R05272.  
 DT 15-AUG-1990. (first entry)  
 DE Polypeptide with amino peptidase-P activity encoded by new gene  
 KW Amino peptidase-P.  
 RN J02002373-A.  
 BD 08-JAN-1990.  
 PF 25-MAR-1989; 071138.  
 PR 25-MAR-1989; JP-071138, JP-156193.  
 PA (AJIN) Ajinomoto KK.  
 PI  
 DR N-PSDB: Q91838.  
 PT Amino peptidase-P-coding gene -  
 PT used in gene-provided recombinant DNA and recombinant  
 PS DNA-provided survival cell stock  
 PS Disclosure: 15pp: Japanese.  
 CC It is new. Also new are recombinant DNA contg. its encoding DNA, cells transformed with the recombinant DNA, and prodn. of it by culturing the cells. The method allows economical, high yielding prodn. of it. It is also useful in separating or refining the enzyme.  
 SO Sequence 441 AA.  
 Query Match  
 Best Local Similarity 66.7%; Score 46; DB 1; Length 441;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Db 305 yrpqtsl1 312  
 |||||  
 OY 2 YRPTVAL 9  
 RESULT  
 ID R97245 standard; Protein; 4472 AA.  
 AC R97245.  
 DT 07-JAN-1997. (first entry)  
 DE Virulence gene cluster polypeptide product.  
 KW Mutant; adaptation; virulence factor; identification; screening; vaccine; drugs; infection; treatment.  
 OS Salmonella typhimurium.

FH Key  
 FT Region  
 FT Location/Qualifiers  
 FT /note- "All x's in this sequence correspond to  
 FT termination codons in the virulence gene  
 FT cluster sequence given in T09224."  
 FT  
 PD W09617951-A2.  
 PD 13-JUN-1996.  
 PF 11-DEC-1995; G02875.  
 PR 09-DEC-1994; GB-024921.  
 PR 31-JAN-1995; GB-001881.  
 PR 05-MAY-1995; GB-009239.  
 PA (RPMS-) RPMS TECHNOLOGY LTD.  
 PI Holden DW.  
 DR WPI: 96-287194/29.  
 PT Identifying virulence genes in microorganisms - by introducing  
 PT mutants with insertion inactivated genes into environment and  
 PT retrieval and analysis of mutants  
 PS Claim 51; Figure 11; 131pp: English.  
 CC A method for identifying a microorganism having a reduced adaptation  
 CC to a particular environment comprising the steps of: (1) providing a  
 CC plurality of microorganisms each of which is independently mutated by  
 CC the insertional inactivation of a gene with a nucleic acid comprising  
 CC a unique marker sequence so that each mutant contains a different  
 CC marker sequence, or clones of the said microorganism; (2) providing  
 CC individually a stored sample of each mutant produced by step (1) and  
 CC providing individually stored nucleic acid comprising the unique  
 CC marker sequence from each individual mutant; (3) introducing a  
 CC plurality of mutants produced by step (1) into the said particular  
 CC environment and allowing those microorganisms which are able to do so  
 CC to grow in the said environment; (4) retrieving microorganisms from  
 CC the said environment or a selected part thereof and isolating the  
 CC nucleic acid from the retrieved microorganisms; (5) comparing any  
 CC unique marker sequence of each individual mutant stored as in step  
 CC (2); and (6) selecting an individual mutant which does not contain any  
 CC of the marker sequences as isolated in step (4). The products and  
 CC methods can be used for identifying virulence genes in microorganisms.  
 CC The mutant microorganisms can be used in vaccines or to screen for  
 CC drugs which reduce virulence or compounds useful for preventing,  
 CC ameliorating or treating infections in animals or plants.  
 SO Sequence 4472 AA.  
 Query Match  
 Best Local Similarity 66.7%; Score 46; DB 20; Length 4472;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 627 yrtgsvgl 634  
 |||||  
 OY 2 YRPTVAL 9  
 RESULT  
 ID R21422 standard; Protein; 31 AA.  
 AC R21422.  
 DT 17-JUN-1992. (first entry)  
 DE Matrix peptide from bovine teeth.  
 KW Insoluble; adjuvant for tissue culture; osteoporosis; growth.  
 OS Bos taurus.  
 PN J04021700-A.  
 PD 24-JAN-1992.  
 PF 16-MAY-1990; 124214.  
 PR 16-MAY-1990; JP-124214.  
 PA (NIPK) NIPPON KAIYAKU KK.  
 DR WPI: 92-076793/10.  
 PT Matrix protein isolated from bovine tooth and bone - useful for  
 PT diagnostic reagent or adjuvant for tissue culture  
 PS Claim 3; Page 1; 6pp: Japanese.  
 CC The matrix peptide was obtd. by isolation from bovine tooth with  
 CC final fractionation by HA-UROGEL chromatography. The peptide has  
 CC Mw 23-25 kD and is insoluble in water, 6M urea and 0.5M guanidine  
 CC hydrochloride. The peptide is useful as a diagnostic reagent or an  
 CC adjuvant for tissue culture. Osteoporosis patients can be diagnosed  
 CC by determining the amt. of the matrix protein antigen in the

CC patients blood by enzyme immunoassay using anti-matrix protein  
 CC antibody. The matrix protein accelerates cell growth in an  
 CC osteoblast culture on a tissue culture plate coated with the matrix  
 CC protein. The matrix protein alone does not exhibit osteosts  
 CC ossification activity or osteoblast differentiation promoting  
 CC activity, but stimulates an osteoblast differentiation activating  
 CC component to exhibit distinct osteosts ossification activity.  
 CC See also R21421-5.  
 SQ Sequence 31 AA;

Query Match 65.2%; Score 45; DB 4; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 1.35e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9 RYRPG 13  
 |||||  
 1 RYRPG 5

RESULT 10  
 ID R12876 standard; Protein; 52 AA.  
 AC R12876;  
 DT 18-SEP-1991 (first entry)  
 DE Non-collagenous bone matrix protein derived S3S1S2 protein.  
 KW Fusion protein; osteoblast; S1; S2; S3.  
 FH Key Location/Qualifiers  
 FT region 1..7 /label= S3  
 FT region 8..35 /label= S1  
 FT region 36..52 /label= S2  
 PN MO9108749-A.  
 PD 27-JUN-1991.  
 PR 20-DEC-1990; E02295.  
 PR 20-DEC-1989; EP-403592.  
 PA (INNO-) INNOGENETICS NV SA.  
 PI Mahy P, Van Heuverswyn H;  
 PI MPI: 91-207858/28.  
 DR Prep. of protein from non-collagenous bone matrix protein - used  
 PT for treatment and diagnosis of Paget's disease and osteosarcoma  
 PS Claim 11; Page 68; 83pp; English.  
 CC The 27 kD protein was purified from an extract of non-collagenous  
 CC bone matrix protein. It inhibits osteoblast proliferation and may  
 CC be used to treat Paget's disease and osteosarcomas. S1, S2, S3 and  
 CC combinations of these may be used diagnostically.  
 CC See also R12872-77.  
 SQ Sequence 52 AA;

Query Match 65.2%; Score 45; DB 3; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.35e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 RYRPG 11  
 |||||  
 1 RYRPG 5

RESULT 11  
 ID R12875 standard; Protein; 52 AA.  
 AC R12875;  
 DT 18-SEP-1991 (first entry)  
 DE Non-collagenous bone matrix protein derived S2S3S1 protein.  
 KW Fusion protein; osteoblast; S1; S2; S3.  
 FH Key Location/Qualifiers  
 FT region 1..16 /label= S2  
 FT region 17..25 /label= S3  
 FT region 26..52 /label= S1  
 PN MO9108749-A.  
 PD 27-JUN-1991.

PF 20-DEC-1990; E02295.  
 PR 20-DEC-1989; EP-403592.  
 PA (INNO-) INNOGENETICS NV SA.  
 PI Mahy P, Van Heuverswyn H;  
 PI MPI: 91-207858/28.  
 DR Prep. of protein from non-collagenous bone matrix protein - used  
 PT for treatment and diagnosis of Paget's disease and osteo:sarcoma  
 PS Claim 11; Page 68; 83pp; English.  
 CC The 27 kD protein was purified from an extract of non-collagenous  
 CC bone matrix protein. It inhibits osteoblast proliferation and may  
 CC be used to treat Paget's disease and osteosarcomas. S1, S2, S3 and  
 CC combinations of these may be used diagnostically.  
 CC See also R12872-77.  
 SQ Sequence 52 AA;

Query Match 65.2%; Score 45; DB 3; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.35e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 23 RYRPG 27  
 |||||  
 1 RYRPG 5

RESULT 12  
 ID R89423 standard; Protein; 239 AA.  
 AC R89423;  
 DT 24-APR-1996 (first entry)  
 DE Mucin-derived protein MUC1/V/alt.  
 KW MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;  
 KW MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;  
 OS receptor; diagnosis; imaging; therapy.  
 OS Homo sapiens.  
 PN MO9603502-A2.  
 PD 08-FEB-1996.  
 PR 21-JUL-1995; IB0627.  
 PR 26-JUL-1994; IL-110464.  
 PA (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.  
 PI Wreschner DH;  
 PI MPI: 96-117047/12.  
 DR N-PSDB; T10682.  
 PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,  
 PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast  
 PT cancer  
 PS Claim 6; Fig 6D; 79pp; English.  
 CC Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,  
 CC MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (R89418-23) are highly and  
 CC differentially expressed in human breast cancer tissue and significantly  
 CC enhance the in vivo tumorigenic potential of mammary epithelial cells.  
 CC They serve as cell surface receptor molecules participating in signal  
 CC transduction. The proteins can be obtained by expression of encoding cDNA  
 CC (see T10677-82) in recombinant host cells. They are used in the  
 CC treatment of human breast cancer and as diagnostic reagents. Receptor  
 CC ligands MUC1/W, W/alt, Z and Z/alt (R89424-27) have also been isolated.  
 SQ Sequence 239 AA;

Query Match 65.2%; Score 45; DB 16; Length 239;  
 Best Local Similarity 44.4%; Pred. No. 1.35e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 77 KFRPGSVV 85  
 ::|||:|:  
 1 RYRPGTVAL 9

RESULT 13  
 ID R89422 standard; Protein; 240 AA.  
 AC R89422;  
 DT 24-APR-1996 (first entry)  
 DE Mucin-derived protein MUC1/V.  
 KW MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;  
 KW MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;  
 KW receptor; diagnosis; imaging; therapy.

OS Homo sapiens.  
 PN WO9603502-A2.  
 PD 08-FEB-1996.  
 PF 21-JUL-1995; IB0627.  
 PR 26-JUL-1994; IL-110464.  
 PA (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.  
 PI Wreschner DH;  
 DR WPI: 96-117047/12.  
 DR N-PSDB: T10681.  
 PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,  
 Z, Z/alt - for the diagnosis, imaging and therapy of human breast  
 cancer.  
 PS Claim 6; Fig 6C; 79pp; English.  
 CC Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,  
 MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (R89418-23) are highly and  
 differentially expressed in human breast cancer tissue and significantly  
 enhance the in vivo tumorigenic potential of mammary epithelial cells.  
 CC They serve as cell surface receptor molecules participating in signal  
 transduction. The proteins can be obtd. by expression of encoding CDNA  
 (see T10677-82) in recombinant host cells. They are used in the  
 CC treatment of human breast cancer and as diagnostic reagents. Receptor  
 CC ligands MUC1/W, W/alt, Z and Z/alt (R89424-27) have also been isolated.  
 SO Sequence 240 AA.

Query Match 65.2%; Score 45; DB 16; Length 240;  
 Best Local Similarity 44.4%; Pred. No. 1.35e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 kfrpgsvvv 86  
 OY 1 RYRPGTVAL 9

RESULT 14  
 ID R89420 standard; Protein: 255 AA.  
 AC R89420;  
 DT 24-APR-1996 (first entry)  
 DE Mucin-derived protein MUC1/X.  
 KM MUC1; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;  
 KW MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;  
 KW receptor; diagnosis; imaging; therapy.  
 OS Homo sapiens.  
 PN WO9603502-A2.  
 PD 08-FEB-1996.  
 PF 21-JUL-1995; IB0627.  
 PR 26-JUL-1994; IL-110464.  
 PA (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.  
 PI Wreschner DH;  
 DR WPI: 96-117047/12.  
 DR N-PSDB: T10679.  
 PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,  
 Z, Z/alt - for the diagnosis, imaging and therapy of human breast  
 cancer.  
 PS Claim 5; Fig 6A; 79pp; English.  
 CC Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,  
 MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (R89418-23) are highly and  
 differentially expressed in human breast cancer tissue and significantly  
 enhance the in vivo tumorigenic potential of mammary epithelial cells.  
 CC They serve as cell surface receptor molecules participating in signal  
 transduction. The proteins can be obtd. by expression of encoding CDNA  
 (see T10677-82) in recombinant host cells. They are used in the  
 CC treatment of human breast cancer and as diagnostic reagents. Receptor  
 CC ligands MUC1/W, W/alt, Z and Z/alt (R89424-27) have also been isolated.  
 SO Sequence 255 AA.

Query Match 65.2%; Score 45; DB 16; Length 255;  
 Best Local Similarity 44.4%; Pred. No. 1.35e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 93 kfrpgsvvv 101  
 OY 1 RYRPGTVAL 9

RESULT 15  
 ID R89421 standard; Protein: 264 AA.  
 AC R89421;  
 DT 24-APR-1996 (first entry)  
 DE Mucin-derived protein MUC1/Y/alt.  
 KM MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;  
 KW MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;  
 KW receptor; diagnosis; imaging; therapy.  
 OS Homo sapiens.  
 PN WO9603502-A2.  
 PD 08-FEB-1996.  
 PF 21-JUL-1995; IB0627.  
 PR 26-JUL-1994; IL-110464.  
 PA (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.  
 PI Wreschner DH;  
 DR WPI: 96-117047/12.  
 DR N-PSDB: T10680.  
 PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,  
 Z, Z/alt - for the diagnosis, imaging and therapy of human breast  
 cancer.  
 PS Claim 5; Fig 6B; 79pp; English.  
 CC Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,  
 MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (R89418-23) are highly and  
 differentially expressed in human breast cancer tissue and significantly  
 enhance the in vivo tumorigenic potential of mammary epithelial cells.  
 CC They serve as cell surface receptor molecules participating in signal  
 transduction. The proteins can be obtd. by expression of encoding CDNA  
 (see T10677-82) in recombinant host cells. They are used in the  
 CC treatment of human breast cancer and as diagnostic reagents. Receptor  
 CC ligands MUC1/W, W/alt, Z and Z/alt (R89424-27) have also been isolated.  
 SO Sequence 264 AA.

Query Match 65.2%; Score 45; DB 16; Length 264;  
 Best Local Similarity 44.4%; Pred. No. 1.35e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 102 kfrpgsvvv 110  
 OY 1 RYRPGTVAL 9

Search completed: Fri Sep 11 13:17:39 1998  
 Job time : 15 secs.

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
 WIRENET  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSrch\_p protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Sep 11 13:17:57 1998; MasPar time 3.34 Seconds  
 Tabular output not generated. 98.470 Million cell updates/sec

Title: >US-08-452-843-11  
 Description: (1-9) from US08452843.pep  
 Perfect Score: 69  
 Sequence: 1 RYRPTVAL 9

Scoring table: PAM 150  
 Gap 15

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: 1:pir1 2:pir2 3:pir3 4:pir4 5:nrl3d

Statistics: Mean 23.177; Variance 29.132; scale 0.796

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No.                        |
|------------|-------|-------------|--------|----|-------------|----------------------------------|
| 1          | 69    | 100.0       | 48     | 2  | B61286      | histone H3 - sandpape 4.34e-04   |
| 2          | 69    | 100.0       | 48     | 2  | A61286      | histone H3 - sea urch 4.34e-04   |
| 3          | 69    | 100.0       | 48     | 2  | A61286      | histone H3 - cycad (E 4.34e-04   |
| 4          | 69    | 100.0       | 60     | 2  | S51664      | histone H3 variant H3 4.34e-04   |
| 5          | 69    | 100.0       | 62     | 2  | B38309      | histone H3.2 - alfalf 4.34e-04   |
| 6          | 69    | 100.0       | 83     | 2  | S42066      | histone H3.3 - Leprot 4.34e-04   |
| 7          | 69    | 100.0       | 119    | 2  | S04521      | histone H3 (clone PH3 4.34e-04   |
| 8          | 69    | 100.0       | 121    | 2  | A02630      | histone H3 - fruit fl 4.34e-04   |
| 9          | 69    | 100.0       | 135    | 1  | HSX131      | histone H3.1 - Africa 4.34e-04   |
| 10         | 69    | 100.0       | 135    | 1  | S59592      | histone H3 (clone CH- 4.34e-04   |
| 11         | 69    | 100.0       | 135    | 1  | HSB03       | histone H3 - bovine 4.34e-04     |
| 12         | 69    | 100.0       | 135    | 2  | S59581      | histone H3 (clones CH 4.34e-04   |
| 13         | 69    | 100.0       | 135    | 2  | S00940      | histone H3 - Volvox C 4.34e-04   |
| 14         | 69    | 100.0       | 135    | 1  | HSX132      | histone H3.2 - Africa 4.34e-04   |
| 15         | 69    | 100.0       | 135    | 1  | HSX132      | histone H3.2 - smallmou 4.34e-04 |
| 16         | 69    | 100.0       | 136    | 2  | A56580      | histone H3 - midge (C 4.34e-04   |
| 17         | 69    | 100.0       | 136    | 2  | S09655      | histone H3 - fruit fl 4.34e-04   |
| 18         | 69    | 100.0       | 136    | 3  | A45941      | histone H3 - Atlantic 4.34e-04   |
| 19         | 69    | 100.0       | 136    | 1  | HSCH3       | histone H3 - chicken 4.34e-04    |
| 20         | 69    | 100.0       | 136    | 2  | I49395      | histone H3.2 protein 4.34e-04    |
| 21         | 69    | 100.0       | 136    | 2  | I49398      | histone H3.1 protein 4.34e-04    |
| 22         | 69    | 100.0       | 136    | 2  | S10097      | histone H3 - fruit fl 4.34e-04   |
| 23         | 69    | 100.0       | 136    | 1  | HSKW3       | histone H3 - Caenorha 4.34e-04   |

| RESULT | ENTRY | ENTRY TITLE | ORGANISM | DATE | Accessions | REFERENCE             | #authors | #journal | #title | Accessions | REFERENCE | #authors | #journal | #title | Accessions | REFERENCE | #authors | #journal | #title |
|--------|-------|-------------|----------|------|------------|-----------------------|----------|----------|--------|------------|-----------|----------|----------|--------|------------|-----------|----------|----------|--------|
| 24     | 69    | 100.0       | 136      | 2    | S32638     | histone H3.1 - Africa | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 25     | 69    | 100.0       | 136      | 2    | I50244     | histone H3.3A - chick | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 26     | 69    | 100.0       | 136      | 2    | S24346     | histone H3.3-like pro | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 27     | 69    | 100.0       | 136      | 2    | S61220     | histone H3.3 - fruit  | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 28     | 69    | 100.0       | 136      | 2    | I57019     | H3 histone - rat      | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 29     | 69    | 100.0       | 136      | 2    | A56618     | histone H3 - spoonwor | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 30     | 69    | 100.0       | 136      | 2    | S57473     | histone H3 - human    | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 31     | 69    | 100.0       | 136      | 2    | JN0687     | histone H3 - sea squi | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 32     | 69    | 100.0       | 136      | 2    | S06743     | histone H3 - mouse    | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 33     | 69    | 100.0       | 136      | 2    | S10168     | histone H3.3A - rabbi | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 34     | 69    | 100.0       | 136      | 2    | S50140     | H3.3 histone - sea ur | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 35     | 69    | 100.0       | 136      | 2    | S34185     | histone H3 - rat      | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 36     | 69    | 100.0       | 136      | 2    | I50245     | histone H3.3B - chick | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 37     | 69    | 100.0       | 136      | 2    | S11315     | histone H3 - polychae | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 38     | 69    | 100.0       | 136      | 2    | I51448     | histone H3 - African  | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 39     | 69    | 100.0       | 136      | 2    | A56654     | histone H3 - Tigrispu | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 40     | 69    | 100.0       | 136      | 2    | S61218     | histone H3.3 - fruit  | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 41     | 69    | 100.0       | 136      | 2    | S01198     | histone H3 - starfish | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 42     | 69    | 100.0       | 136      | 2    | I49397     | histone H3.2 protein  | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 43     | 69    | 100.0       | 136      | 2    | S20669     | histone H3 - starfish | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 44     | 69    | 100.0       | 136      | 1    | HSU033     | histone H3.3 - human  | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |
| 45     | 69    | 100.0       | 136      | 2    | I50460     | H3 histone - muscovy  | 4.34e-04 |          |        |            |           |          |          |        |            |           |          |          |        |

## ALIGNMENTS

RESULT 1  
 ENTRY B61286 #type fragment  
 histone H3 - sandpape limpet (fragment)  
 ORGANISM #formal name Patella granatina #common name sandpape limpet  
 DATE 12-May-1994 #sequence-revision 12-May-1994 #text-change 03-May-1996

ACCESSIONS  
 B61286  
 REFERENCE  
 Brandt, W.F.; Strickland, W.N.; Morgan, M.; Von Holt, C.  
 FEBS Lett. (1974) 40:167-172  
 #title Comparison of the N-terminal amino acid sequences of histone F3 from a mammal, a bird, a shark, an echinoderm, a mollusc and a plant.

#accession B61286 preliminary  
 #status preliminary  
 #molecule-type protein  
 #residues 1-48 #label BRA  
 CLASSIFICATION #superfamily histone H3  
 SUMMARY #length 48 #checksum 171

Query Match 100.0%; Score 69; DB 2; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 4.34e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPTVAL 48  
 |||||  
 Oy 1 RYRPTVAL 9

RESULT 2  
 ENTRY A61286 #type fragment  
 histone H3 - sea urchin (Parechlinus angulosus) (fragment)  
 ORGANISM #formal name Parechlinus angulosus #common name angulate urchin  
 DATE 12-May-1994 #sequence-revision 12-May-1994 #text-change 03-May-1996

ACCESSIONS  
 A61286  
 REFERENCE  
 Brandt, W.F.; Strickland, W.N.; Morgan, M.; Von Holt, C.  
 FEBS Lett. (1974) 40:167-172  
 #title Comparison of the N-terminal amino acid sequences of histone F3 from a mammal, a bird, a shark, an echinoderm, a mollusc and a plant.

#accession A61286 preliminary  
 #status preliminary  
 #molecule-type protein  
 #residues 1-48 #label BRA

CLASSIFICATION #superfamily histone H3  
SUMMARY #length 48 #checksum 171

Query Match 100.0%; Score 69; DB 2; Length 48;  
Best Local Similarity 100.0%; Pred. No. 4.34e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
1 RYRPGTVAL 9

RESULT 3

ENTRY histone H3 - cycad (Encephalartos caffer) (fragment)  
TITLE histone H3 - cycad (Encephalartos caffer) (fragment)  
ORGANISM #formal\_name Encephalartos caffer #common\_name cycad  
DATE 19-Mar-1997 #sequence\_revision 19-Dec-1997 #text\_change 26-Feb-1998

ACCESSIONS C61286  
REFERENCE A61286

#authors Brandt, W.F.; Strickland, W.N.; Morgan, M.; Von Holt, C.  
#journal FEBS Lett. (1974) 40:167-172  
#title Comparison of the N-terminal amino acid sequences of histone F3 from a mammal, a bird, a shark, an echinoderm, a mollusc and a plant.

#accession C61286  
#status preliminary  
#molecule\_type protein  
#residues 1-48 #label BRA  
CLASSIFICATION #superfamily histone H3  
SUMMARY #length 48 #checksum 171

Query Match 100.0%; Score 69; DB 2; Length 48;  
Best Local Similarity 100.0%; Pred. No. 4.34e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
1 RYRPGTVAL 9

RESULT 4

ENTRY histone H3 variant H3.3 - tomato (fragment)  
TITLE histone H3 variant H3.3 - tomato (fragment)  
ORGANISM #formal\_name Lycopersicon esculentum #common\_name tomato  
DATE 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 03-May-1996

ACCESSIONS S51664  
REFERENCE S51664

#authors Hartung, F.  
#journal submitted to the EMBL Data Library, December 1994  
#accession S51664  
#status preliminary  
#molecule\_type mRNA  
#residues 1-60 #label HAR  
CLASSIFICATION #cross-references EMBL:X83422  
SUMMARY #superfamily histone H3  
#length 60 #checksum 8500

Query Match 100.0%; Score 69; DB 2; Length 60;  
Best Local Similarity 100.0%; Pred. No. 4.34e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 41 RYRPGTVAL 49  
1 RYRPGTVAL 9

RESULT 5

ENTRY histone H3.2 - alfalfa (fragments)  
TITLE histone H3.2 - alfalfa (fragments)  
ORGANISM #formal\_name Medicago sativa #common\_name alfalfa  
DATE 14-Jun-1991 #sequence\_revision 14-Jun-1991 #text\_change

ACCESSIONS 23-Feb-1997  
REFERENCE B38309  
#authors Waterborg, J.H.  
#journal J. Biol. Chem. (1990) 265:17157-17161  
#title Sequence analysis of acetylation and methylation in two histone H3 variants of alfalfa.

#cross-references MVID:91009145  
#accession B38309  
#status preliminary  
#molecule\_type protein  
#residues 1-62 #label WAT  
CLASSIFICATION #superfamily histone H3  
KEYWORDS chromosomal protein; DNA binding; nucleosome core; nucleus  
SUMMARY #length 62 #checksum 7971

Query Match 100.0%; Score 69; DB 2; Length 62;  
Best Local Similarity 100.0%; Pred. No. 4.34e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
1 RYRPGTVAL 9

RESULT 6

ENTRY histone H3.3 - Leptothorax acervorum (fragment)  
TITLE histone H3.3 - Leptothorax acervorum (fragment)  
ORGANISM #formal\_name Leptothorax acervorum  
DATE 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 13-Mar-1998

ACCESSIONS S42066  
REFERENCE S42066

#authors Baur, A.; Stetzer, N.E.; Buschinger, A.; Zimmermann, F.K.  
#journal Submitted to the EMBL Data Library, February 1994  
#description Cloning of two differentially expressed reverse transcription fragments of the histone 3 gene of Leptothorax acervorum (Hymenoptera Formicidae).

#accession S42066  
#molecule\_type mRNA  
#residues 1-83 #label BAU  
#cross-references EMBL:X77741; NID:9456196; PID:9456197  
#experimental\_source freshly laid eggs  
#accession S42067  
#molecule\_type mRNA  
#residues 1-83 #label BAW  
#cross-references EMBL:X77740  
#experimental\_source larvae, pupae and adults

GENETICS H3.3

CLASSIFICATION #superfamily histone H3  
KEYWORDS chromosomal protein; DNA binding; nucleosome core; nucleus  
SUMMARY #length 83 #checksum 1158

Query Match 100.0%; Score 69; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 4.34e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 8 RYRPGTVAL 16  
1 RYRPGTVAL 9

RESULT 7

ENTRY histone H3 (clone PH3C-1) - alfalfa (fragment)  
TITLE histone H3 (clone PH3C-1) - alfalfa (fragment)  
ORGANISM #formal\_name Medicago sativa #common\_name alfalfa  
DATE 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 08-Sep-1997

ACCESSIONS S04521  
REFERENCE S04520

#authors Wu, S.C.; Gyoergye, J.; Dudits, D.  
#journal Nucleic Acids Res. (1989) 17:3057-3063

```
#title Polyadenylated H3 histone transcripts and H3 histone variants
#cross-references MUID:89263717
#accession S04521
##molecule-type mRNA
##residues 1-119 ##label WUS
CLASSIFICATION #superfamily histone H3
KEYWORDS DNA binding; nucleus
SUMMARY #length 119 #checksum 3419

# Query Match 100.0%; Score 69; DB 2; Length 119;
# Best Local Similarity 100.0%; Pred. No. 4.34e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 24 RYRPGTVAL 32
QY 1 RYRPGTVAL 9

RESULT 8
ENTRY A02630 #type fragments
TITLE histone H3 - fruit fly (Drosophila melanogaster) (fragments)
ORGANISM #formal_name Drosophila melanogaster
DATE 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change
23-Feb-1997
A02630
ACCESSION A02630
REFERENCE Goldberg, M.L.
#authors Ph.D. thesis, Stanford Univ., 1979
#accession A02630
##molecule-type DNA
##residues 1-121 ##label GOL
#note the author translated the codon CCC for residue 31 as
Ala

GENETICS
#gene FlyBase:H33
#cross-references FlyBase:FBgn0001199
CLASSIFICATION #superfamily histone H3
KEYWORDS chromosomal protein; DNA binding; nucleosome core; nucleus
SUMMARY #length 121 #checksum 3455

# Query Match 100.0%; Score 69; DB 2; Length 121;
# Best Local Similarity 100.0%; Pred. No. 4.34e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48
QY 1 RYRPGTVAL 9

RESULT 9
ENTRY H5XL31 #type complete
TITLE histone H3.1 - African clawed frog
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change
20-Mar-1998
A93596
ACCESSION A93596; A02634; A24279; E24510
REFERENCE Old, R.W.; Sheikh, S.A.; Chambers, A.; Newton, C.A.;
#authors Mohammed, A.; Aldridge, T.C.
#journal Nucleic Acids Res. (1985) 13:7341-7558
#title Individual Xenopus histone genes are replication-independent
in oocytes and replication-dependent in Xenopus or mouse
somatic cells.
#cross-references MUID:86041919
#accession A93596
##molecule-type DNA
##residues 1-135 ##label OLD
#cross-references GB:X03104; NID:964780; PID:964781
#experimental_source clone XLHW23
REFERENCE A92918
#authors Perry, M.; Thomsen, G.H.; Roeder, R.G.
```

```
#journal J. Mol. Biol. (1985) 185:479-499
#title Genomic organization and nucleotide sequence of two distinct
histone gene clusters from Xenopus laevis. Identification
of novel conserved upstream sequence elements.
#cross-references MUID:86037224
#accession A92918
##molecule-type DNA
##residues 1-135 ##label PER
COMMENT #experimental_source clones Xh1 and Xh3
The initiator Met is not shown.
CLASSIFICATION #superfamily histone H3
KEYWORDS chromosomal protein; DNA binding; nucleosome core
SUMMARY #length 135 #molecular-weight 15257 #checksum 8291

# Query Match 100.0%; Score 69; DB 1; Length 135;
# Best Local Similarity 100.0%; Pred. No. 4.34e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48
QY 1 RYRPGTVAL 9

RESULT 10
ENTRY S59592 #type complete
TITLE histone H3 (clone CH-1) - Chlamydomonas reinhardtii
ORGANISM #formal_name Chlamydomonas reinhardtii
DATE 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change
08-Sep-1997
S59592
ACCESSION S59581
REFERENCE Fabry, S.; Mueller, K.; Lindauer, A.; Park, P.B.; Cornelius,
#authors T.; Schmitt, R.
#journal Curr. Genet. (1995) 28:333-345
#title The organization structure and regulatory elements of
Chlamydomonas histone genes reveal features linking plant
and animal genes.
#accession S59592
##status preliminary
##molecule-type DNA
##residues 1-135 ##label FAB
#cross-references EMBL:U16825; NID:9576632; PID:9576633
#note the authors did not translate the codon for residue 1
sequence not shown
CLASSIFICATION #superfamily histone H3
SUMMARY #length 135 #molecular-weight 15296 #checksum 7516

# Query Match 100.0%; Score 69; DB 2; Length 135;
# Best Local Similarity 100.0%; Pred. No. 4.34e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48
QY 1 RYRPGTVAL 9

RESULT 11
ENTRY HSB03 #type complete
TITLE histone H3 - bovine
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 08-Oct-1981 #sequence_revision 08-Oct-1981 #text_change
16-Feb-1997
A02624; PLO130; A49978
ACCESSION A02624
REFERENCE Delange, R.J.; Hooper, J.A.; Smith, E.L.
#authors J. Biol. Chem. (1973) 248:3261-3274
#journal Histone III. III. Sequence studies on the cyanogen bromide
peptides; complete amino acid sequence of calf thymus
histone III.
#cross-references MUID:73166574
#accession A02624
##molecule-type protein
##residues 1-135 ##label DEL
```

#cross-references MUID:88234003



```

#accession      S00940
#molecule_type DNA
#residues      1-135 #label MUE
#cross-references EMBL:X06963; NID:g21983; PID:g21985
GENETICS
#introns      46/1
CLASSIFICATION #superfamily histone H3
KEYWORDS      chromosomal protein; DNA binding; nucleosome core; nucleus
SUMMARY      #length 135 #molecular-weight 15310 #checksum 7476

Query Match      100.0%; Score 69; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 4.34e-04;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      40 RYRPGTVAL 48
QY      1 RYRPGTVAL 9

RESULT      14
ENTRY      HSYL32      #type complete
TITLE      histone H3.2 African clawed frog
ORGANISM   #formal_name Xenopus laevis #common_name African clawed frog
DATE      18-Aug-1982 #sequence_revision 18-Aug-1982 #text_change
16-Feb-1997
A02634
A91300
ACCESSIONS #authors      Moorman, A.F.M.; De Boer, P.A.J.; De Laaf, R.T.M.; Van
REFERENCE   Dongen, W.M.A.M.; Destree, O.H.J.
#journal    FEBS Lett. (1981) 136:45-52
#title      Primary structure of the histone H3 and H4 genes and their
            flanking sequences in a minor histone gene cluster of
            Xenopus laevis.
#cross-references NID:8205633
#accession   A02634
#molecule_type DNA
#residues    1-135 #label MOR
#experimental_source clone Xlhl
#note        #initiator Met not shown
CLASSIFICATION #superfamily histone H3
KEYWORDS      chromosomal protein; DNA binding; nucleosome core
SUMMARY      #length 135 #molecular-weight 15356 #checksum 7803

Query Match      100.0%; Score 69; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 4.34e-04;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      40 RYRPGTVAL 48
QY      1 RYRPGTVAL 9

RESULT      15
ENTRY      HSF13
TITLE      histone H3 - smallmouth buffalo fish
ORGANISM   #formal_name Ictiobus bubalus #common_name smallmouth buffalo
            fish
DATE      30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change
16-Feb-1997
A02627
A02627
ACCESSIONS #authors      Hooper, J.A.; Smith, E.L.; Sommer, K.R.; Chalkley, R.
REFERENCE   J. Biol. Chem. (1973) 248:3275-3279
#journal    Histone III. IV. Amino acid sequence of histone III of the
            testes of the carp, Ictiobus bubalus.
#cross-references MUID:7316575
#accession   A02627
#molecule_type protein
#residues    1-135 #label HCO
#note        Lys-9 is epsilon-N-monomethyllysine,
            epsilon-N-dimethyllysine, epsilon-N-trimethyllysine,
            or unmodified in 10, 13, 11, and 65% of the molecules,
            respectively

```

```

#note
Lys-27 is epsilon-N-monomethyllysine,
epsilon-N-dimethyllysine, epsilon-N-trimethyllysine,
or unmodified in 36, 40, 16, and 8% of the molecules,
respectively
CLASSIFICATION #superfamily histone H3
KEYWORDS      chromosomal protein; DNA binding; methylated amino acid;
            nucleosome core
FEATURE      9,27
#modified_site N6-methyllysine, N6,N6-dimethyllysine or
            N6,N6-trimethyllysine (Lys) (partial) #status
            experimental
SUMMARY      #length 135 #molecular-weight 15257 #checksum 8291

Query Match      100.0%; Score 69; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 4.34e-04;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      40 RYRPGTVAL 48
QY      1 RYRPGTVAL 9

```

Search completed: Fri Sep 11 13:18:09 1998  
Job time : 12 secs.

**THIS PAGE BLANK (USPTO)**



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSearch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:18:26 1998; Maspar time 2.39 Seconds

Tabular output not generated. 94.293 Million cell updates/sec

Title: >US-08-452-843-11  
Description: (1-9) from US08452843.pep  
Perfect Score: 69  
Sequence: 1 RYRGTVAL 9

Scoring table: PAM 150  
Gap 15

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: swiss-prot35  
1:swiss1

Statistics: Mean 24.149; Variance 25.248; scale 0.956

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID        | Description            | Pred. No. |
|------------|-------|-------------|--------|----|-----------|------------------------|-----------|
| 1          | 69    | 100.0       | 134    | 1  | H3_VOLCA  | HISTONE H3.            | 2.37e-05  |
| 2          | 69    | 100.0       | 134    | 1  | H3_CHLBE  | HISTONE H3.            | 2.37e-05  |
| 3          | 69    | 100.0       | 135    | 1  | H3_CAEEL  | HISTONE H3.            | 2.37e-05  |
| 4          | 69    | 100.0       | 135    | 1  | H3_DROME  | HISTONE H3.            | 2.37e-05  |
| 5          | 69    | 100.0       | 135    | 1  | H32_MEDSA | HISTONE H3.2, MINOR.   | 2.37e-05  |
| 6          | 69    | 100.0       | 135    | 1  | H32_BOVIN | HISTONE H3 (H3.2).     | 2.37e-05  |
| 7          | 69    | 100.0       | 135    | 1  | H31_SCHPO | HISTONE H3.1/H3.2.     | 2.37e-05  |
| 8          | 69    | 100.0       | 135    | 1  | H31_HUMAN | HISTONE H3.1.          | 2.37e-05  |
| 9          | 69    | 100.0       | 135    | 1  | H3_PSAHI  | HISTONE H3.            | 2.37e-05  |
| 10         | 69    | 100.0       | 135    | 1  | H3_ACRFO  | HISTONE H3.            | 2.37e-05  |
| 11         | 69    | 100.0       | 135    | 1  | H33_HUMAN | HISTONE H3.            | 2.37e-05  |
| 12         | 69    | 100.0       | 135    | 1  | H33_CAEEL | HISTONE H3.3 (H3.B) (H | 2.37e-05  |
| 13         | 69    | 100.0       | 135    | 1  | H32_XENLA | HISTONE H3.2.          | 2.37e-05  |
| 14         | 69    | 100.0       | 135    | 1  | H3_STRPU  | HISTONE H3.2.          | 2.37e-05  |
| 15         | 69    | 100.0       | 135    | 1  | H31_TETPY | HISTONE H3.1.          | 2.37e-05  |
| 16         | 69    | 100.0       | 135    | 1  | H32_ORYSA | HISTONE H3.            | 2.37e-05  |
| 17         | 69    | 100.0       | 135    | 1  | H3_PEA    | HISTONE H3.            | 2.37e-05  |
| 18         | 69    | 100.0       | 135    | 1  | H3_NEUCR  | HISTONE H3.            | 2.37e-05  |
| 19         | 69    | 100.0       | 135    | 1  | H3_TEAET  | HISTONE H3.            | 2.37e-05  |
| 20         | 69    | 100.0       | 135    | 1  | H3_EMENT  | HISTONE H3.            | 2.37e-05  |
| 21         | 69    | 100.0       | 135    | 1  | H3_MAIZE  | HISTONE H3.            | 2.37e-05  |
| 22         | 69    | 100.0       | 135    | 1  | H3_ENCAL  | HISTONE H3.            | 2.37e-05  |
| 23         | 69    | 100.0       | 134    | 1  | H34_MOUSE | HISTONE H3.4 (EMBRYONI | 1.90e-03  |

|    |    |      |      |   |            |                         |          |
|----|----|------|------|---|------------|-------------------------|----------|
| 24 | 60 | 87.0 | 135  | 1 | H33_SCHPO  | HISTONE H3.3.           | 6.36e-03 |
| 25 | 60 | 87.0 | 135  | 1 | H34_CAIHO  | HISTONE H3.4.           | 6.36e-03 |
| 26 | 59 | 85.5 | 135  | 1 | H33_TETTH  | HISTONE H3.3 (HV2).     | 1.15e-02 |
| 27 | 59 | 85.5 | 135  | 1 | H32_TETPY  | HISTONE H3.2.           | 1.15e-02 |
| 28 | 56 | 81.2 | 413  | 1 | NCAP_HNV   | NUCLEOCAPSID PROTEIN (  | 6.62e-02 |
| 29 | 56 | 81.2 | 530  | 1 | KPYL_FELCA | PYRUVATE KINASE, M1 (M  | 6.62e-02 |
| 30 | 54 | 78.3 | 125  | 1 | ULAG_HCMVA | HYPOTHETICAL PROTEIN U  | 2.06e-01 |
| 31 | 53 | 76.8 | 274  | 1 | PSBO_SYNY3 | PHOTOSYSTEM II MANGANE  | 3.59e-01 |
| 32 | 51 | 73.9 | 261  | 1 | YLB2_CAEEL | HYPOTHETICAL HISTONE 3  | 1.07e+00 |
| 33 | 51 | 73.9 | 288  | 1 | YMB3_CAEEL | HYPOTHETICAL HISTONE 3  | 1.07e+00 |
| 34 | 51 | 73.9 | 344  | 1 | POLG_MCFE  | GENOME POLYPROTEIN (CO  | 1.07e+00 |
| 35 | 50 | 72.5 | 505  | 1 | YML8_YEAST | HYPOTHETICAL 57.7 KD P  | 1.82e+00 |
| 36 | 50 | 72.5 | 682  | 1 | VG50_BPM15 | GENE 50 PROTEIN (GP50)  | 1.82e+00 |
| 37 | 49 | 71.0 | 213  | 1 | YKB2_CAEEL | HYPOTHETICAL 24.7 KD P  | 3.08e+00 |
| 38 | 49 | 71.0 | 294  | 1 | POL_SNSAV  | POL POLYPROTEIN (REVER  | 3.08e+00 |
| 39 | 49 | 71.0 | 527  | 1 | KPKR_XENLA | ATP-CITRATE (PRO-S-)-L  | 3.08e+00 |
| 40 | 49 | 71.0 | 1100 | 1 | ACLY_RAT   | ATP-CITRATE (PRO-S-)-L  | 3.08e+00 |
| 41 | 49 | 71.0 | 1106 | 1 | ACLY_CAEEL | ATP-CITRATE (PRO-S-)-L  | 3.08e+00 |
| 42 | 48 | 69.6 | 63   | 1 | YH14_RHOCA | PROBABLE ATP-CITRATE (  | 3.08e+00 |
| 43 | 48 | 69.6 | 354  | 1 | Y48K_CAMJE | HYPOTHETICAL PROTEIN I  | 5.16e+00 |
| 44 | 48 | 69.6 | 440  | 1 | Y48K_CAMJE | FRUCTOSE-BISPHOSPHATE   | 5.16e+00 |
| 45 | 48 | 69.6 | 794  | 1 | SC18_CANAL | HYPOTHETICAL 48 KD PRO  | 5.16e+00 |
|    |    |      |      |   |            | VESTITULAR-FUSION PROTE | 5.16e+00 |

## ALIGNMENTS

| RESULT                | 1   | STANDARD: | PRT: | 134 AA. |
|-----------------------|---|-----------|------|---------|
| ID                    | H3_VOLCA  |           |      |         |
| AC                    | P08437:   |           |      |         |
| DT                    | 01-AUG-1988 (REL. 08, CREATED)  |           |      |         |
| DT                    | 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)                           |           |      |         |
| DT                    | 01-AUG-1990 (REL. 15, LAST ANNOTATION UPDATE)                         |           |      |         |
| DE                    | HISTONE H3.   |           |      |         |
| GN                    | H3-I AND H3-II.   |           |      |         |
| OS                    | VOLVOX CARTERI.   |           |      |         |
| OC                    | EUKARYOTA; PLANTA; PHYCOPHYTA; CHLOROPHYTA (GREEN ALGAE);             |           |      |         |
| OC                    | CHLOROPHYCEAE; VOLVOCALES; VOLVOCAECAE.                               |           |      |         |
| RN                    | [1]   |           |      |         |
| RP                    | SEQUENCE FROM N.A.  |           |      |         |
| RC                    | STRAIN-HK10;  |           |      |         |
| RX                    | MEDLINE; 88234003.  |           |      |         |
| RA                    | MUELLER K., SCHMITT R.;   |           |      |         |
| RL                    | NUCLEIC ACIDS RES. 16:4121-4136(1988).                                |           |      |         |
| CC                    | -I- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE |           |      |         |
| CC                    | IN NUCLEOSOME FORMATION.  |           |      |         |
| CC                    | -I- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF |           |      |         |
| CC                    | H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.       |           |      |         |
| DR                    | EMBL; X06963; G21985; -   |           |      |         |
| DR                    | EMBL; X06964; G21988; -   |           |      |         |
| DR                    | PIR; S00940; S00940.  |           |      |         |
| DR                    | PROSITE; PS00322; HISTONE_H3_1.                                       |           |      |         |
| DR                    | PROSITE; PS00959; HISTONE_H3_2.                                       |           |      |         |
| DR                    | NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.   |           |      |         |
| FT                    | INIT MET 0  |           |      |         |
| FT                    | SEQUENCE 134 AA; 15179 MW; 0FE12547 CRC32;                            |           |      |         |
| Query Match           | 100.0%; Score 69; DB 1; Length 134;                                   |           |      |         |
| Best Local Similarity | 100.0%; Pred. No. 2.37e-05;   |           |      |         |
| Matches               | 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;                    |           |      |         |
| DB                    | 39 RYRGTVAL 47  |           |      |         |
| OY                    | 1 RYRGTVAL 9  |           |      |         |
| RESULT                | 2   | STANDARD: | PRT: | 134 AA. |
| ID                    | H3_CHLBE  |           |      |         |
| AC                    | P50564:   |           |      |         |
| DT                    | 01-OCT-1996 (REL. 34, CREATED)  |           |      |         |
| DT                    | 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)                           |           |      |         |
| DT                    | 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)                         |           |      |         |
| DE                    | HISTONE H3.   |           |      |         |
| GN                    | H3.   |           |      |         |

OS CHLAMYDOMONAS REINHARDTII.  
 CC EUKARYOTA: PLANTA: PHYCOPHYTA: CHLOROPHYTA (GREEN ALGAE);  
 CC CHLOROPHYCEAE: VOLVOCALES: CHLAMYDOMONADACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-137.  
 RA MEDLINE: 96017782.  
 RL WALTER 2., HALL J.L.;  
 CC NUCLEIC ACIDS RES. 23:3756-3763(1995).  
 CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
 IN NUCLEOSOME FORMATION.  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
 DR EMBL: L41841; G790700;  
 DR PROSITE: PS00323; HISTONE\_H3\_1; 1.  
 DR PROSITE: PS00959; HISTONE\_H3\_2; 1.  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.  
 FT INIT\_MET 0  
 FT SEQUENCE 134 AA; 15129 MW; 281606B6 CRC32;  
 SQ  
 Query Match 100.0%; Score 69; DB 1; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 2.37e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 39 RYRPGTVAL 47  
 1 RYRPGTVAL 9

RESULT 3  
 ID H3-CAEEL STANDARD: PRT: 135 AA.  
 AC P08898;  
 DT 01-NOV-1988 (REL. 09, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE HISTONE H3;  
 GN (HIS-9 OR F22B3.2) AND F55G1.2 AND B0035.10.  
 OS CAENORHABDITIS ELEGANS.  
 CC EUKARYOTA: METAEOA: ACCELLOMATES; NEMATODA: SECERNENTEA; RHABDITIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE: 89293823.  
 RA ROBERTS S.B., EMMONS S.W., CHILDS G.;  
 RL J. MOL. BIOL. 206:567-577(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A. (F22B3.2).  
 RC STRAIN-BRISTOL N2;  
 RA COTTAGE A.;  
 RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A. (F55G1.2).  
 RC STRAIN-BRISTOL N2;  
 RA WATERSTON R.;  
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [4]  
 RP SEQUENCE FROM N.A. (B0035.10).  
 RC STRAIN-BRISTOL N2;  
 RA WHITE S.;  
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [5]  
 RP SEQUENCE.  
 RC STRAIN-DR27;  
 RX MEDLINE: 87105951.  
 RA VANETTEREN J.R., VAN BUN S.M., VAN BEEMEN J.J.;  
 RL FEBS LETT. 211:59-63(1987).  
 CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
 IN NUCLEOSOME FORMATION.  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
 CC EMBL: X16341; G6752;  
 DR EMBL: 268336; E217365;  
 DR EMBL: U58750; G1326373;  
 DR EMBL: Z73102; E242587;

DR PIR: A25842; HSKW3.  
 DR PIR: S04241; S04241.  
 DR WORMPEP: B0035.10; CE03253.  
 DR WORMPEP: F22B3.2; CE03253.  
 DR WORMPEP: F55G1.2; CE03253.  
 DR PROSITE: PS00322; HISTONE\_H3\_1; 1.  
 DR PROSITE: PS00959; HISTONE\_H3\_2; 1.  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE;  
 FT ACETYLATION; METHYLATION.  
 FT INIT\_MET 0  
 FT SEQUENCE 135 AA; 15213 MW; 43A8A478 CRC32;  
 SQ  
 Query Match 100.0%; Score 69; DB 1; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 2.37e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
 1 RYRPGTVAL 9

RESULT 4  
 ID H3-DROME STANDARD: PRT: 135 AA.  
 AC P02299;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)  
 DE HISTONE H3.  
 GN HIS3.  
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
 CC EUKARYOTA: METAEOA: ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MATSUO Y., YAMAZAKI T.;  
 RL NUCLEIC ACIDS RES. 17:225-238(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA GOLDBERG M.L.;  
 RL THEISIS (1979), UNIVERSITY OF STANFORD, U.S.A.  
 CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
 IN NUCLEOSOME FORMATION.  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
 CC EMBL: X14215; G8070;  
 DR PIR: A02630; A02630.  
 DR FLYBASE: FBgn0001199; HIS3.  
 DR PROSITE: PS00322; HISTONE\_H3\_1; 1.  
 DR PROSITE: PS00959; HISTONE\_H3\_2; 1.  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE;  
 FT ACETYLATION; METHYLATION.  
 FT INIT\_MET 0  
 FT SEQUENCE 135 AA; 15285 MW; C3050F48 CRC32;  
 SQ  
 Query Match 100.0%; Score 69; DB 1; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 2.37e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
 1 RYRPGTVAL 9

Sun Sep 13 10:55:14 1998

QY 1 RYRPGTVAL.9

RESULT 5 STANDARD: PRT: 135 AA.

ID H32\_MEDSA

AC P1105;

DT 01-JUL-1989 (REL. 11, CREATED)

DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)

DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE HISTONE H3.2, MINOR.

OS MEDICAGO SATIVA (ALFALFA), ARABIDOPSIS THALIANA (MOUSE-EAR CRESS), AND LOTIUM TEMULENTUM.

OC EUKARYOTA: PLANTA: EMBRYOPHYTA: ANGIOSPERMAE: DICOTYLEDONEAE: FABALES: OC FABACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=M.SATIYA; STRAIN=CV. CHIEF, AND CV. REGEN S;

RA ROBERTSON A.J.;

RL THESES (1994), UNIVERSITY OF MISSOURI-KANSAS CITY, U.S.A.

RN [2]

RP SEQUENCE OF 17-135 FROM N.A.

RC SPECIES=M.SATIYA; STRAIN=CV. REGEN S;

RA MEDLINE: 89263717.

RA WU S.C., GYOERGEY J., DUDITS D.;

RL NUCLEIC ACIDS RES. 17:3057-3063(1989).

RN [3]

RP SEQUENCE OF 1-44 AND 84-115, ACETYLATION, AND METHYLATION.

RC SPECIES=M.SATIYA; STRAIN=CV. R4;

RA MEDLINE: 91009145.

RA WATERBORG J.H.;

RL J. BIOL. CHEM. 265:17157-17161(1990).

RN [4]

RP SEQUENCE FROM N.A.

RC SPECIES=A.THALIANA; STRAIN=CV. COLUMBIA;

RA MEDLINE: 92277663.

RA CHAUBET N., CLEMENT B., GIGOT C.;

RL J. MOL. BIOL. 225:569-574(1992).

RN [5]

RP SEQUENCE FROM N.A.

RC SPECIES=L.TEMULENTUM;

RA FREEMAN D.R., OUGHAM H.J.;

RL SUBMITTED (JUL-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE IN NUCLEOSOME FORMATION.

CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.

CC CC

CC EMBL: U09458; G488563; -

CC DR EMBL: U09460; G488567; -

CC DR EMBL: U09461; G488569; -

CC DR EMBL: U09464; G488575; -

CC DR EMBL: U09465; G488577; -

CC DR EMBL: X13676; G829279; -

CC DR EMBL: X60429; G404825; -

CC DR EMBL: X60429; G16324; -

CC DR EMBL: X79714; G510911; -

CC DR PIR: B38309; B38309;

CC DR PIR: S04521; S04521;

CC DR PIR: S24346; S24346;

CC DR PROSITE: PS00322; HISTONE\_H3.1; 1.

CC DR PROSITE: PS00959; HISTONE\_H3.2; 1.

CC KM NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE; MULTIGENE FAMILY; ACETYLATION; METHYLATION.

FT INIT MET 0 0

FT MOD\_RES 4 4 METHYLATION.

FT MOD\_RES 9 9 METHYLATION.

FT MOD\_RES 14 14 ACETYLATION.

FT MOD\_RES 14 14 ACETYLATION.

FT MOD\_RES 18 18 ACETYLATION.

FT MOD\_RES 18 18 ACETYLATION.

FT MOD\_RES 18 18 ACETYLATION.

FT MOD\_RES 23 23 ACETYLATION.

FT MOD\_RES 23 23 METHYLATION.

FT MOD\_RES 27 27 METHYLATION.

SO SEQUENCE 135 AA; 15275 MW; F32F962A CRC32;

Query Match 100.0%; Score 69; DB 1: Length 135;

Best Local Similarity 100.0%; Pred. No. 2.37e-05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48

QY 1 RYRPGTVAL 9

RESULT 6 STANDARD: PRT: 135 AA.

ID H32\_BOVIN

AC P16105; P02295; P02297; P17320; P17269;

DT 21-JUL-1986 (REL. 01, CREATED)

DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)

DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)

DE HISTONE H3 (H3.2).

OS BOS TAURUS (BOVINE), MUS MUSCULUS (MOUSE), GALLUS GALLUS (CHICKEN), CAIRINA MOSCHATA (MUSCOVY DUCK), ICTIOBUS BUBALUS (SMALLMOUTH BUFFALO FISH), PORODERA AFRICANUS (AFRICAN CATFISH) (AFRICAN CATSHARK), XENODUS LAEYIS (AFRICAN CLAWED FROG), ONCORHYNCHUS MYKISS (RAINBOW TROUT) (SALMO GAIARDNERI), PLATYNERIS DUMERILII (DUMERIL'S CLAM WORM), OS DROSOPHILA HYDEI (FRUIT FLY), TIGRIPODUS CALIFORNICUS, CHIRONOMUS THUMMI THUMMI (MIDGE), AND URECHIS CAUPO (INNKEEPER WORM) (SPOONWORM).

OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA.

RN [1]

RP SEQUENCE.

RC SPECIES=BOVINE; TISSUE=THYMUS;

RA MEDLINE: 72154496.

RA MARZUFF W.F. JR., SANDERS L.A., MILLER D.M., MCCARTY K.S.;

RL J. BIOL. CHEM. 247:2026-2033(1972).

RN [2]

RP SEQUENCE.

RC SPECIES=BOVINE;

RA MEDLINE: 75095680.

RA PATHY L., SMITH E.L.;

RL J. BIOL. CHEM. 250:1919-1920(1975).

RN [3]

RP SEQUENCE FROM N.A.

RC SPECIES=MOUSE;

RA MEDLINE: 84041477.

RA SITMAN D.B., GAYLES R.A., MARZUFF W.F.;

RL NUCLEIC ACIDS RES. 11:6679-6697(1983).

RN [4]

RP SEQUENCE FROM N.A.

RC SPECIES=MOUSE; STRAIN=BALB/C; TISSUE=LIVER;

RA MEDLINE: 90067871.

RA HURT M.M., CHODCHOY N., MARZUFF W.F.;

RL NUCLEIC ACIDS RES. 17:8876-8876(1989).

RN [5]

RP SEQUENCE FROM N.A. (H3.2-221 AND H3.2-614).

RC SPECIES=MOUSE;

RA MEDLINE: 87112762.

RA TAYLOR J.D., WELLMAN S.E., MARZUFF W.F.;

RL J. MOL. EVOL. 23:242-249(1986).

RN [6]

RP SEQUENCE FROM N.A.

RC SPECIES=MOUSE;

RA MEDLINE: 91065547.

RA GRUER A., STREIT A., REIST M., BENNINGER P., BOEHNI R., SCHENKPERL D.;

RL GENE 95:303-304(1990).

RN [7]

RP SEQUENCE FROM N.A.

RC SPECIES=CHICKEN;

RA MEDLINE: 85215552.

RA WANG S.W., ROBINS A.J., D'ANDREA R., WELLS J.R.E.;

RL NUCLEIC ACIDS RES. 13:1369-1387(1985).

RN [8]

RP SEQUENCE FROM N.A.

RC SPECIES=CHICKEN;

RA MEDLINE: 82195575.

RA ENGEL J.D., SUGARMAN B.J., DODGSON J.B.;  
 RL NATURE 297:434-436(1982).  
 RN [19]  
 RP SEQUENCE FROM N.A. (H3-II AND H3-III).  
 RC SPECIES-CHICKEN;  
 RA NAKAYAMA T.;  
 RL GENE 102:289-290(1991).  
 RN [10]  
 RP SEQUENCE FROM N.A. (H3-IV AND H3-V).  
 RC SPECIES-CHICKEN; STRAIN-WHITE LEGHORN;  
 RA SETOGUCHI Y., NAKAYAMA T.;  
 RL NUCLEIC ACIDS RES. 19:6327-6327(1991).  
 RN [11]  
 RP SEQUENCE.  
 RC SPECIES-CHICKEN;  
 RA BRANDT W.F., VON HOLT C.;  
 RL EUR. J. BIOCHEM. 46:419-429(1974).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-DUCK;  
 RX MEDLINE: 89178747.  
 RA TOENJES R., MONK K., DOENECKE D.;  
 RL J. MOL. EVOL. 28:200-211(1989).  
 RN [13]  
 RP SEQUENCE.  
 RC SPECIES-I. BUBALUS;  
 RX MEDLINE: 7316575.  
 RA HOOVER J.A., SMITH E.L., SOMMER K.R., CHALKLEY R.;  
 RL J. BIOL. CHEM. 248:3275-3279(1973).  
 RN [14]  
 RP SEQUENCE.  
 RC SPECIES-P. AFRICANUS;  
 RX MEDLINE: 74309063.  
 RA BRANDT W.F., STRICKLAND W.N., VON HOLT C.;  
 RL FEBS LETT. 40:349-352(1974).  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-O. MYKISS;  
 RX MEDLINE: 85083109.  
 RA CONNOR W., STATES J.C., MEZQUITA J., DIXON G.H.;  
 RL J. MOL. EVOL. 20:236-250(1984).  
 RN [16]  
 RP SEQUENCE OF 1-25.  
 RC SPECIES-O. MYKISS;  
 RX MEDLINE: 72259090.  
 RA CANDIDO E.P.M., DIXON G.H.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 69:2015-2019(1972).  
 RN [17]  
 RP SEQUENCE FROM N.A. (CLONE XLHW23).  
 RC SPECIES-X. LAEYIS;  
 RX MEDLINE: 86041919.  
 RA OLD R.W., SHEIKH S.A., CHAMBERS A., NEWTON C.A., MOHAMMED A.,  
 RL ALDRIDGE T.C.;  
 RN NUCLEIC ACIDS RES. 13:7341-7358(1985).  
 RN [18]  
 RP SEQUENCE FROM N.A. (GENE CLUSTERS X1H1 AND X1H3).  
 RC SPECIES-X. LAEYIS;  
 RX MEDLINE: 86037224.  
 RA PERRY M., THOMSEN G.H., ROEDER R.G.;  
 RL J. MOL. BIOL. 185:479-499(1985).  
 RN [19]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-P. DUMERILLI; TISSUE-SPERM;  
 RX MEDLINE: 90306006.  
 RA SELLOS D., KRAMERZ S.A., DIXON G.H.;  
 RL EUR. J. BIOCHEM. 190:21-29(1990).  
 RN [20]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-D. HYDEI;  
 RX MEDLINE: 90221886.

RA KREMER H., HENNIG W.;  
 RL NUCLEIC ACIDS RES. 18:1573-1580(1990).  
 RN [21]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-D. HYDEI;  
 RA STRAUSBAUGH L.D., FITCH D.H.A., BARRETT V.;  
 RL SUBMITTED (APR-1990) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [22]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-T. CALIFORNICUS;  
 RX MEDLINE: 92127050.  
 RA PORTER D., BROWN D., WELLS D.;  
 RL DNA SEQ. 1:197-206(1991).  
 RN [23]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-T. CALIFORNICUS;  
 RX MEDLINE: 93076000.  
 RA BROWN D., COOK A., WAGNER M., WELLS D.;  
 RL DNA SEQ. 2:387-396(1992).  
 RN [24]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-C.T. THUMMI;  
 RX MEDLINE: 91039309.  
 RA HANKELN T., SCHMIDT E.R.;  
 RL J. MOL. BIOL. 215:477-482(1990).  
 RN [25]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-C.T. PIGER;  
 RX MEDLINE: 94087747.  
 RA HANKELN T., SCHMIDT E.R.;  
 RL J. MOL. BIOL. 234:1301-1307(1993).  
 RN [26]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-U. CAUPO; TISSUE-SPERM;  
 RX MEDLINE: 92329960.  
 RA DAVIS F.C., SHELTON J.C., INGHAM L.D.;  
 RL DNA SEQ. 2:247-256(1992).  
 CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
 CC IN NUCLEOSOME FORMATION.  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.  
 CC  
 DR EMBL: X01685; G51319; -;  
 DR EMBL: X16148; G51301; -;  
 DR EMBL: M32459; G387195; -;  
 DR EMBL: M32461; G387197; -;  
 DR EMBL: M33989; G387183; -;  
 DR EMBL: X02218; G63474; -;  
 DR EMBL: J00869; G211855; ALT\_SEQ.  
 DR EMBL: M61154; G211857; -;  
 DR EMBL: M61155; G211859; -;  
 DR EMBL: X62291; G63482; -;  
 DR EMBL: X62292; G63484; -;  
 DR EMBL: X14732; G62735; -;  
 DR EMBL: X14732; G62734; -;  
 DR EMBL: X01064; G64326; -;  
 DR EMBL: X03104; G64781; -;  
 DR EMBL: X03017; G64772; -;  
 DR EMBL: X03018; G64778; -;  
 DR EMBL: X53330; G9821; -;  
 DR EMBL: X17072; G7439; -;  
 DR EMBL: X52576; G7433; -;  
 DR EMBL: X52393; G10615; -;  
 DR EMBL: M84797; G161895; -;  
 DR EMBL: X56335; G7083; -;  
 DR EMBL: X72803; G297563; -;

\*\*\* Note: remainder of annotations omitted.

Query Match 100.0%; Score 69; DB 1; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 2.37e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48

OY 1 RYRPGTVAL 9  
 RESULT 7  
 ID H31\_SCHPO STANDARD: PRT; 135 AA.  
 AC P09988;  
 DT 01-MAR-1989 (REL. 10, CREATED)  
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
 DE HISTONE H3.1/H3.2  
 GM HHT1 AND HHT2  
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 86135992.  
 RL MATSUMOTO S., YANAGIDA M.;  
 EMO J. 4:3531-3538(1985).  
 CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
 IN NUCLEOSOME FORMATION.  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.  
 DR EMBL: X05222; G4970; -;  
 DR EMBL: X05223; G4963; -;  
 DR PIR: E27399; HSZP3.  
 DR PROSITE: PS00322; HISTONE\_H3\_1; 1.  
 DR PROSITE: PS00959; HISTONE\_H3\_2; 1.  
 KM NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE;  
 MULTIGENE FAMILY.  
 KW INIT\_MET  
 FT SEQUENCE 135 AA; 15226 MW; 9932E953 CRC32;  
 SQ  
 Query Match 100.0%; Score 69; DB 1; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 2.37e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 40 RYRPGTVAL 48  
 OY 1 RYRPGTVAL 9  
 RESULT 8  
 ID H31\_HUMAN STANDARD: PRT; 135 AA.  
 AC P16106; P02295; P02296;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)  
 DE HISTONE H3.1.  
 OS HOMO SAPIENS (HUMAN). BOS TAURUS (BOVINE), AND MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX SPECIES-HUMAN;  
 MEDLINE: 84069776.  
 RA ZHONG R., ROEDER R.G., HEINTZ N.;  
 NUCLEIC ACIDS RES. 11:7409-7425(1983).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC SPECIES-HUMAN;  
 MEDLINE: 86242753.  
 RA MARASHI F., HELMS S., SHIELDS A., SILVERSTEIN S., GREENSPAN D.S.,  
 STEIN G., STEIN J.;  
 BIOCHEM. CELL BIOL. 64:277-289(1986).  
 RL [3]  
 RN SEQUENCE FROM N.A.  
 RC SPECIES-HUMAN;  
 MEDLINE: 92009931.  
 RA ALBIG W., KARALINOU E., DRABENT B., ZIMMER A., DOENECKE D.;  
 GENOMICS 10:940-948(1991).  
 RL [4]  
 RN PARTIAL SEQUENCE.

RC SPECIES-HUMAN;  
 RX MEDLINE: 82075746.  
 RA OHE Y., IWAI K.;  
 RL J. BIOCHEM. 90:1205-1211(1981).  
 RN [5]  
 RP SEQUENCE.  
 RC SPECIES-BOVINE;  
 RX MEDLINE: 73166574.  
 RA DELANGE R.J., HOOPER J.A., SMITH E.L.;  
 RL J. BIOL. CHEM. 248:3261-3274(1973).  
 RN [6]  
 RP PARTIAL SEQUENCE.  
 RC SPECIES-BOVINE;  
 RX MEDLINE: 73166572.  
 RA DELANGE R.J., SMITH E.L.;  
 RL J. BIOL. CHEM. 248:3248-3254(1973).  
 RN [7]  
 RP PARTIAL SEQUENCE.  
 RC SPECIES-BOVINE;  
 RX MEDLINE: 73166573.  
 RA HOOPER J.A., SMITH E.L.;  
 RL J. BIOL. CHEM. 248:3255-3260(1973).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-MOUSE; STRAIN-CD-1; TISSUE-TESTIS;  
 RX MEDLINE: 90067856.  
 RA KOSCIENSA U., DOENECKE D.;  
 RL NUCLEIC ACIDS RES. 17:8861-8861(1989).  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-MOUSE;  
 RX MEDLINE: 84041477.  
 RA SITTMAN D.B., GRAVES R.A., MARZLUFF W.F.;  
 RL NUCLEIC ACIDS RES. 11:6679-6697(1983).  
 RN [10]  
 RP SEQUENCE FROM N.A. (H3.1-221 AND H3.1-291).  
 RC SPECIES-MOUSE;  
 RX MEDLINE: 87112762.  
 RA TAYLOR J.D., WELLMAN S.E., MARZLUFF W.F.;  
 RL J. MOL. EVOL. 23:242-249(1986).  
 CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
 IN NUCLEOSOME FORMATION.  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.  
 CC EMBL: X00090; G32115; -;  
 DR EMBL: M26150; G386772; -;  
 DR EMBL: M60746; G184070; -;  
 DR EMBL: X57128; G31982; -;  
 DR EMBL: X16496; G51327; -;  
 DR EMBL: X01684; G51317; -;  
 DR EMBL: M32460; G387196; -;  
 DR EMBL: M32462; G387198; -;  
 DR PIR: A02623; HSHU3.  
 DR PIR: A02624; HSHO3.  
 DR PIR: A40335; A40335.  
 DR PIR: S06755; S06755.  
 DR PIR: S28528; S28528.  
 DR MM: 142780; -;  
 DR PROSITE: PS00322; HISTONE\_H3\_1; 1.  
 DR PROSITE: PS00959; HISTONE\_H3\_2; 1.  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE;  
 MULTIGENE FAMILY; ACETYLATION; METHYLATION.  
 RN INIT\_MET 0  
 FT MOD\_RES 9 9 METHYLATION.  
 FT MOD\_RES 14 14 ACETYLATION.  
 FT MOD\_RES 23 23 ACETYLATION.  
 FT MOD\_RES 27 27 METHYLATION.  
 FT MOD\_RES 36 36 METHYLATION.  
 FT CONFLICT 134 134 MISSING (IN REF. 2).  
 SQ SEQUENCE 135 AA; 15273 MW; BC3247C6 CRC32;  
 Query Match 100.0%; Score 69; DB 1; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 2.37e-05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
 |||||||  
 1 RYRPGTVAL 9

RESULT 9  
 ID H3\_PSAMI STANDARD: PRT: 135 AA.  
 AC P02298: P05320: P05321: P05322:  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HISTONE H3, EMBRYONIC.  
 OS PSAMECHINUS MILIARIS (SAND SEA URCHIN), PARACENTRUS LIVIDUS (COMMON SEA URCHIN), STRONGILOCENTROTUS DROBACHIIENSIS (SEA URCHIN),  
 OS DEMASTERIAS IMBRICATA (SEA STAR), PISASTER BREVISPIRUS (SEA STAR),  
 OS PISASTER OCHRACEUS (SEA STAR), PYCNOPODIA HELIANTHOIDES (SEA STAR),  
 OS AND SOLASTER STIMPSONI (SEA STAR).  
 CC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;  
 CC EUCECHINOIDEA.  
 RN [1]  
 RP SEQUENCE FROM N.A. (CLONE H22).  
 RC SPECIES=P.MILIARIS;  
 RX MEDLINE: 79001915.  
 RA SCHAFFNER W., KUNZ G., DAETWYLER H., TELFORD J., SMITH H.O.,  
 RA BIRNSTIEL M.L.;  
 RL CELL 14:655-671(1978).  
 RN [2]  
 RP SEQUENCE FROM N.A. (CLONE H22).  
 RC SPECIES=P.MILIARIS;  
 RA BIRNSTIEL M.L., PORTMANN R., BUSSLINGER M., SCHAFFNER W., PROBST E.,  
 RA KRESSMANN A.;  
 RL ALFRED BENZON SYMP. 13:117-132(1979).  
 RN [3]  
 RP SEQUENCE FROM N.A. (CLONE H19).  
 RC SPECIES=P.MILIARIS;  
 RX MEDLINE: 81076674.  
 RA BUSSLINGER M., PORTMANN R., IRMINGER J.C., BIRNSTIEL M.L.;  
 RA NUCLEIC ACIDS RES. 8:957-977(1980).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=P.LIVIDUS;  
 RX SPINELLI G.;  
 RL SUBMITTED (FEB-1989) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=P.LIVIDUS, AND S.DROBACHIIENSIS;  
 RA BUSSLINGER M., RUSCONI S., BIRNSTIEL M.L.;  
 RL EMO J. 1:27-33(1982).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=P.IMBRICATA, P.BREVISPIRUS, AND P.OCHRACEUS; TISSUE=SPERM;  
 RX MEDLINE: 88259237.  
 RA BANFIELD D.C.D., HONDA B.M., SMITH M.J.;  
 RL J. MOL. EVOL. 27:36-44(1988).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=P.OCHRACEUS, P.BREVISPIRUS, P.HELLANTHOIDES, AND S.STIMPSONI;  
 RA WU Y., KOMBEL D., SMITH M.J.;  
 RL SUBMITTED (JUL-1990) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
 IN NUCLEOSOME FORMATION.  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 H2A, H2B, H3, AND H4, WHICH WRAP APPROXIMATELY 146 BP OF DNA.  
 CC -1- DEVELOPMENTAL STAGE: THIS HISTONE IS EXPRESSED DURING LATE  
 EMBRYONIC DEVELOPMENT.  
 CC EMBL: J01181, G161401; -;  
 DR EMBL: X01345, G10035; -;  
 DR EMBL: V01143, G10021; -;  
 DR EMBL: V01144, E15894; ALT\_SEQ.  
 DR EMBL: M10558, G16187; -;  
 DR EMBL: M25281, G159970; -;

DR EMBL: X07505; G7468; -;  
 DR EMBL: M36919; G159867; -;  
 DR EMBL: M36920; G159965; -;  
 DR EMBL: M36921; G16181; -;  
 DR EMBL: X07504; G9792; -;  
 DR EMBL: X07503; G10048; -;  
 DR EMBL: X54112; G9790; -;  
 DR EMBL: X54113; G10046; -;  
 DR EMBL: X54114; G9991; -;  
 DR EMBL: X54115; G10340; -;  
 DR PIR: A02628; HSUR3M.  
 DR PIR: S01196; S01196.  
 DR PIR: S01197; S01197.  
 DR PIR: S01198; S01198.  
 DR PIR: S20667; S20667.  
 DR PIR: S20671; S20671.  
 DR PIR: S20678; S20678.  
 DR PIR: S20669; S20669.  
 DR PROSITE: P500322; HISTONE\_H3\_1; 1.  
 DR PROSITE: P500959; HISTONE\_H3\_2; 1.  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE;  
 KW EMBRYO.  
 FT INIT\_MET 0  
 SQ SEQUENCE 135 AA; 15271 MW; 6AD6F728 CRC32;

Query Match 100.0%; Score 69; DB 1; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 2.37e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
 |||||||  
 1 RYRPGTVAL 9

RESULT 10  
 ID H3\_ACRFO STANDARD: PRT: 135 AA.  
 AC P22843;  
 DT 01-AUG-1991 (REL. 19, CREATED)  
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE HISTONE H3.  
 OS ACROPORA FORMOSA (STAGHORN CORAL).  
 CC EUKARYOTA; METAZOA; CNIDARIA; ANTHOZOA; ZOANTHARIA; SCLERACTINIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE: 91033046.  
 RA MILLER D.J., MCMILLAN J., MILES A., TEN LOHUIS M., MAHONY T.;  
 RL GENE 93:319-320(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94047119.  
 RA MILLER D.J., HARRISON P.L., MAHONY T.J., MCMILLAN J.P., MILES A.,  
 RA ODORICO D.M., TEN LOHUIS M.R.;  
 RL J. MOL. EVOL. 37:245-253(1993).  
 CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
 IN NUCLEOSOME FORMATION.  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 H2A, H2B, H3, AND H4, WHICH WRAP APPROXIMATELY 146 BP OF DNA.  
 CC EMBL: M60509; G155640; -;  
 DR EMBL: L11067; G166308; -;  
 DR EMBL: S67324; G455649; -;  
 DR PIR: J00757; J00757.  
 DR PROSITE: P500322; HISTONE\_H3\_1; 1.  
 DR PROSITE: P500959; HISTONE\_H3\_2; 1.  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.  
 FT INIT\_MET 0  
 SQ SEQUENCE 135 AA; 15154 MW; 99B89C5B CRC32;

Query Match 100.0%; Score 69; DB 1; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 2.37e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48



Sun Sep 13 10:55:14 1998

OY 1 RYRPGTVAL 9

RESULT 11  
ID H33\_HUMAN STANDARD: PRT; 135 AA.  
AC P06351; P33155;  
DT 01-JAN-1988 (REL. 06, CREATED)  
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HISTONE H3.3 (H3.B) (H3.30).  
GN H3F3B OR H3H3.30 OR H3S.3B.  
OS HOMO SAPIENS (HUMAN), MUS MUSCULUS (MOUSE), RATTUS NORVEGICUS (RAT),  
OS ORYZOCTOLAGUS CUNICULUS (RABBIT), GALLUS GALLUS (CHICKEN), SPISULA  
OS SOLIDISSIMA (ATLANTIC SURF-CLAM), DROSOPHILA MELANOGASTER (FRUIT  
OS FLY), AND DROSOPHILA HYDEI (FRUIT FLY).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES-HUMAN;  
RX MEDLINE: 85190590.  
RA WELLS D., KEDES L.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 82:2834-2838(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES-HUMAN;  
RX MEDLINE: 87174815.  
RA WELLS D., HOFFMAN D., KEDES L.;  
RL NUCLEIC ACIDS RES. 15:2871-2889(1987).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-HUMAN;  
RX MEDLINE: 96163879.  
RA ALBIG W., BRAMBLAGE B., GRUBER K., KLOBECK H.-G., KUNZ J., DOENECKE D.;  
RL GENOMICS 30:264-272(1995).  
RN [4]  
RP PARTIAL SEQUENCE.  
RC SPECIES-HUMAN;  
RX MEDLINE: 82075746.  
RA OHE Y., IMAI K.;  
RL J. BIOCHEM. 90:1205-1211(1981).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES-MOUSE;  
RX MEDLINE: 89240011.  
RA HRABA-RENEVEY S., KRESS M.;  
RL NUCLEIC ACIDS RES. 17:2449-2461(1989).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC SPECIES-RAT; STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;  
RA DI LIEGRO I.;  
RL SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [7]  
RP SEQUENCE FROM N.A.  
RC SPECIES-RABBIT;  
RX MEDLINE: 90272438.  
RA CHALMERS M., WELLS D.;  
RL NUCLEIC ACIDS RES. 18:3075-3075(1990).  
RN [8]  
RP SEQUENCE FROM N.A.  
RC SPECIES-CHICKEN;  
RX MEDLINE: 85295962.  
RA BRUSH D., DODGSON J.B., CHOI O.R., WILKINS STEVENS P., ENGEL J.D.;  
RL MOL. CELL. BIOL. 5:1307-1317(1985).  
RN [9]  
RP SEQUENCE FROM N.A.  
RC SPECIES-CHICKEN; STRAIN-WHITE LEGHORN; TISSUE-LIVER;  
RX MEDLINE: 87316886.  
RA DODGSON J.B., YAMAMOTO M., ENGEL J.D.;  
RL NUCLEIC ACIDS RES. 15:6294-6294(1987).  
RN [10]  
RP SEQUENCE FROM N.A.

RC SPECIES-SOLIDISSIMA;  
RX MEDLINE: 87305176.  
RA SWENSON K.I., BORGESE N., PIETRINI G., RUDERMAN J.V.;  
RL DEV. BIOL. 123:10-16(1987).  
RN [11]  
RP SEQUENCE FROM N.A.  
RC SPECIES-D. MELANOGASTER;  
RX MEDLINE: 92084129.  
RA FRETZIN S., ALLAN B.D., VAN DAAL A., ELGIN S.C.R.;  
RL GENE 107:341-342(1991).  
RN [12]  
RP SEQUENCE FROM N.A.  
RC SPECIES-D. HYDEI; STRAIN-TUBINGEN;  
RX MEDLINE: 96023949.  
RA AKHANOVA A.S., BINDELS P.S.T., XU J., MIEDMA K., KREMER H.,  
RA HENNING W.;  
RL GENOME 38:586-600(1995).  
CC -!- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
CC H2A, H2B, H3, AND H4, WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
CC -!- THIS HISTONE IS THE PREDOMINANT FORM IN NONDIVIDING CELLS.  
DR EMBL: M11354; G306848; -;  
DR EMBL: M11353; G306849; -;  
DR EMBL: 248950; G761716; -;  
DR EMBL: X05855; E219000; ALT\_SEQ.  
DR EMBL: X05856; E219000; JOINED.  
DR EMBL: X05857; E219000; JOINED.  
DR EMBL: X73683; G313319; -;  
DR EMBL: X13605; G51198; -;  
DR EMBL: X51897; G1568; -;  
DR EMBL: M11383; G211853; -;  
DR EMBL: Y00392; G63480; -;  
DR EMBL: M17876; G161190; -;  
DR EMBL: X53822; G8046; -;  
DR EMBL: X82257; G1006654; -;  
DR EMBL: X81205; G963031; -;  
DR EMBL: X81206; G963024; -;  
DR EMBL: X81207; G963029; -;  
DR EMBL: X81208; G963026; -;  
DR PIR: A02622; HSHU33.  
DR PIR: A27501; A27501.  
DR PIR: S04186; S04186.  
DR PIR: S10168; S10168.  
DR PIR: A45941; A45941.  
DR PIR: J01343; J01343.  
DR FLYBASE: FBgn0004828; H1s3.3B.  
DR MIM: 601058; -;  
DR PROSITE: PS00322; HISTONE\_H3\_1; 1.  
DR PROSITE: PS00959; HISTONE\_H3\_2; 1.  
KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE;  
KW MULTIGENE FAMILY.  
FT INIT MET 0  
SQ SEQUENCE 135 AA: 15197 MW: 5880974A CRC32;  
Query Match 100.0%; Score 69; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. No. 2.37e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 40 RYRPGTVAL 48  
OY 1 RYRPGTVAL 9  
RESULT 12  
ID H33\_CABEL STANDARD: PRT; 135 AA.  
AC Q10453;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HISTONE H3.3.  
GN F45EL 6  
OS CAENORHABDITIS ELEGANS.

```

OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTERA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
RA WATERSTON R.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
IN NUCLEOSOME FORMATION.
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
DR EMBL: U28732; G660702; -;
DR WORKPAP: F4581.6; CE1048; -;
DR PROSITE: PS00322; HISTONE_H3_1; 1;
DR PROSITE: PS00959; HISTONE_H3_2; 1;
KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE;
MULTIGENE FAMILY.
FT INT_MET 0
SQ SEQUENCE 135 AA; 15211 MW; 2167D92 CRC32;

Query Match 100.0%; Score 69; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.37e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRGTVAL 48
OY 1 RYRGTVAL 9

RESULT 13
ID H32_XENLA STANDARD; PRT; 135 AA.
AC P02302;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HISTONE H3.2.
OS XENOPUS LAEVIS (AFRICAN CLAMLED FROG).
NC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN [1]
RP SEQUENCE FROM N.A. (GENE CLUSTER X1H1).
RX MEDLINE: 82095633.
RA MOORMAN A.F.M., DE BOER P.A.J., DE LAAF R.T.M., VAN DONGEN W.M.A.M.,
DESTRE O.H.J.;
RL FEBS LETT. 136:45-52(1981).
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
IN NUCLEOSOME FORMATION.
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
DR EMBL: J00982; -; NOT_ANNOTATED_CDS.
DR PIR: A02634; HSX132.
DR PROSITE: PS00322; HISTONE_H3_1; 1;
DR PROSITE: PS00959; HISTONE_H3_2; 1;
KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE;
MULTIGENE FAMILY.
FT INT_MET 0
SQ SEQUENCE 135 AA; 15356 MW; 9DA3E094 CRC32;

Query Match 100.0%; Score 69; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.37e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRGTVAL 48
OY 1 RYRGTVAL 9

RESULT 14
ID H3_STRPU STANDARD; PRT; 135 AA.
AC P06352;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1990 (REL. 15, LAST ANNOTATION UPDATE)
DE HISTONE H3; EMBRYONIC.

```

```

OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86232591.
RA KAUMEYER J.F., WEINBERG E.S.;
RL NUCLEIC ACIDS RES. 14:4557-4576(1986).
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
IN NUCLEOSOME FORMATION.
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
DR EMBL: X03952; G10258; -;
DR PIR: A02629; HSUR3P.
DR PROSITE: PS00322; HISTONE_H3_1; 1;
DR PROSITE: PS00959; HISTONE_H3_2; 1;
KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE;
EMBRYO.
FT INT_MET 0
SQ SEQUENCE 135 AA; 15370 MW; F9EBF865 CRC32;

Query Match 100.0%; Score 69; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.37e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRGTVAL 48
OY 1 RYRGTVAL 9

RESULT 15
ID H31_TETPR STANDARD; PRT; 135 AA.
AC P15311;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HISTONE H3.1.
GN HHT1 AND HHT2.
OS TETRAHYMENA PYRIFORMIS, AND TETRAHYMENA THERMOPHILA.
OC EUKARYOTA; PROTOZOA; CILIOPHORA; CILIATA; HOLOTRICHA; HYMENOSTOMATIDA.
RN [1]
RP SEQUENCE.
RX SPECIES-T. PYRIFORMIS;
RX MEDLINE: 84289353.
RA HAYASHI T., HAYASHI H., FUSAUCHI Y., IWAI K.;
RL J. BIOCHEM. 95:1741-1749(1984).
RN [2]
RP SEQUENCE OF 1-41 FROM N.A.
RX SPECIES-T. PYRIFORMIS;
RX MEDLINE: 90221813.
RA BRUNK C.F., SADLER L.A.;
RL NUCLEIC ACIDS RES. 18:323-329(1990).
RN [3]
RP SEQUENCE OF 1-41 FROM N.A.
RX SPECIES-T. PYRIFORMIS;
RX MEDLINE: 90219078.
RA BRUNK C.F., KAHN R.W., SADLER L.A.;
RL J. MOL. EVOL. 30:290-297(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX SPECIES-T. TETRAHYMENA;
RX MEDLINE: 94167244.
RA THACHER T.H., MACCAFFEE J., BOWEN J., HOROWITZ S., SHAPIRO D.L.,
RA GOROVSKY M.A.;
RL NUCLEIC ACIDS RES. 22:180-186(1994).
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
IN NUCLEOSOME FORMATION.
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
DR EMBL: X17141; G578553; -;
DR EMBL: M87304; G161788; -;

```

DR EMBL: M87504; G161790; -  
 DR PIR: A28852; A28852.  
 DR PIR: S41499; S41499.  
 DR PROSITE: PS00322; HISTONE\_H3\_1; 1.  
 DR PROSITE: PS00959; HISTONE\_H3\_2; 1.  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE;  
 KW MULTIGENE FAMILY.  
 FT: INIT MET 0  
 SQ SEQUENCE 135 AA; 15298 MW; 5AAD5B94 CRC32;

Query Match: 92.88; Score 64; DB 1; Length 135;  
 % Identity 88.98; Pred. NO. 5.57e-04;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Caps 0;

DB 40 RRRPGTVAL 48  
 1:|||||

QY 1 RRRPGTVAL 9

Search completed: Fri Sep 11 13:18:33 1998  
 Job time : 7 secs.

**THIS PAGE BLANK (USPTO)**

(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Fri Sep 11 13:18:50 1998;  MasPar time 3.85 Seconds
```

Tabular output not generated.

```

Title: >US-08-452-843-11
Description: (1-9) from US08452843.pep
Perfect Score: 69
Sequence: 1 RYRPGTVAL 9

```

Scoring table: PAM 150

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

```
Database:      spreml6
1:sp_fungi 2:sp_human 3:sp_invertebrate 4:sp_mammal
5:sp_mhc 6:sp_organelle 7:sp_phage 8:sp_plant
9:sp_bacteria 10:sp_rodent 11:sp_virus 12:sp_vertebrate
13:sp_unclassified
```

Statistics: Mean 23.385; Variance 25.221; scale 0.927

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description           | Pred. No. |
|------------|-------|-------------|--------|----|--------|-----------------------|-----------|
| 1          | 69    | 100.0       | 56     | 8  | 042754 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 2          | 69    | 100.0       | 56     | 8  | 042776 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 3          | 69    | 100.0       | 56     | 8  | 045031 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 4          | 69    | 100.0       | 56     | 8  | 042743 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 5          | 69    | 100.0       | 56     | 8  | 043029 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 6          | 69    | 100.0       | 56     | 8  | 042724 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 7          | 69    | 100.0       | 56     | 8  | 043030 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 8          | 69    | 100.0       | 56     | 8  | 042755 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 9          | 69    | 100.0       | 56     | 8  | 042821 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 10         | 69    | 100.0       | 56     | 8  | 043228 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 11         | 69    | 100.0       | 56     | 8  | 043820 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 12         | 69    | 100.0       | 56     | 8  | 042759 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 13         | 69    | 100.0       | 56     | 8  | 042789 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 14         | 69    | 100.0       | 56     | 8  | 023976 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 15         | 69    | 100.0       | 56     | 8  | 042745 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 16         | 69    | 100.0       | 56     | 8  | 042754 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 17         | 69    | 100.0       | 56     | 8  | 042817 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 18         | 69    | 100.0       | 56     | 8  | 042825 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 19         | 69    | 100.0       | 56     | 8  | 042818 | HISTONE H3 (FRAGMENT) | 3.93e-05  |
| 20         | 69    | 100.0       | 56     | 8  | 042722 | HISTONE H3 (FRAGMENT) | 3.93e-05  |

|    |    |       |     |    |        |                       |          |
|----|----|-------|-----|----|--------|-----------------------|----------|
| 21 | 69 | 100.0 | 56  | 8  | Q43313 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 22 | 69 | 100.0 | 56  | 8  | Q41612 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 23 | 69 | 100.0 | 56  | 8  | Q2758  | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 24 | 69 | 100.0 | 56  | 8  | Q4274  | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 25 | 69 | 100.0 | 56  | 8  | Q4275  | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 26 | 69 | 100.0 | 83  | 3  | Q27674 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 27 | 69 | 100.0 | 88  | 3  | P91949 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 28 | 69 | 100.0 | 90  | 3  | P91947 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 29 | 69 | 100.0 | 124 | 8  | P94020 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 30 | 69 | 100.0 | 127 | 8  | Q43303 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 31 | 69 | 100.0 | 127 | 8  | Q43303 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 32 | 69 | 100.0 | 127 | 8  | Q42782 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 33 | 69 | 100.0 | 127 | 8  | Q42832 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 34 | 69 | 100.0 | 127 | 8  | Q43202 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 35 | 69 | 100.0 | 127 | 8  | P94019 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 36 | 69 | 100.0 | 135 | 8  | Q42681 | HISTONE H3 (FRAGMENT) | 3.93e-05 |
| 37 | 69 | 100.0 | 136 | 12 | Q92068 | HISTONE H3            | 3.93e-05 |
| 38 | 69 | 100.0 | 136 | 12 | Q92153 | HISTONE H3            | 3.93e-05 |
| 39 | 69 | 100.0 | 136 | 3  | Q27689 | HISTONE H3            | 3.93e-05 |
| 40 | 69 | 100.0 | 136 | 3  | Q27718 | HISTONE H3            | 3.93e-05 |
| 41 | 69 | 100.0 | 136 | 3  | Q02648 | HISTONE H3            | 3.93e-05 |
| 42 | 69 | 100.0 | 136 | 3  | Q27866 | HISTONE H3            | 3.93e-05 |
| 43 | 69 | 100.0 | 136 | 2  | Q93081 | HISTONE H3            | 3.93e-05 |
| 44 | 69 | 100.0 | 136 | 10 | Q64528 | HISTONE H3            | 3.93e-05 |
| 45 | 69 | 100.0 | 136 | 3  | Q27719 | HISTONE H3            | 3.93e-05 |

## ALIGNMENTS

|  |   |              |          |                 |
|--|---|--------------|----------|-----------------|
| RESULT   | 1   | PRELIMINARY; | PRT;     | 56 AA.          |
| ID   | Q42754  |              |          |                 |
| AC   | Q42754; Q42751;   |              |          |                 |
| DT   | 01-NOV-1996 (TREMBLREL. 01, CREATED)                                |              |          |                 |
| DT   | 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)                   |              |          |                 |
| DT   | 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)                 |              |          |                 |
| DE   | HISTONE H3 (FRAGMENT).  |              |          |                 |
| OS   | GLYCINE CLANDESTINA.  |              |          |                 |
| OC   | EUKARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;     |              |          |                 |
| OC   | ROSIDAE; FABALES; FABACEAE.   |              |          |                 |
| RN   | [1]   |              |          |                 |
| RP   | SEQUENCE FROM N.A.  |              |          |                 |
| RX   | MEDLINE; 97131650.  |              |          |                 |
| RA   | DOYLE J.T., KANAZIN V., SHOEWAKER R.C.;                             |              |          |                 |
| RL   | MOL. PHYLOGENET. EVOL. 6:438-447(1996).                             |              |          |                 |
| DR   | EMBL; U47380; G1208684; -.  |              |          |                 |
| DR   | EMBL; U47380; G1208650; -.  |              |          |                 |
| KW   | NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE. |              |          |                 |
| FT   | NON_TER   | 1            |          |                 |
| FT   | NON_TER   | 56           | 56       |                 |
| SO   | SEQUENCE  | 56 AA;       | 6506 MW; | AA28B116 CRC32; |
| Query Match  |   |              |          |                 |
| Best Local Similarity 100.0%; Score 69; DB 8; Length 56;   |   |              |          |                 |
| Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |   |              |          |                 |
| DB   | 1 RYRPTVAL 25   |              |          |                 |
| QY   | 1 RYRPTVAL 9  |              |          |                 |

RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
 RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).  
 DR EMBL: U47405; G1208697;  
 DR EMBL: U47369; G1213303;  
 DR EMBL: U47370; G1213305;  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.  
 FT NON\_TER 1 1  
 SO SEQUENCE 56 AA: 6506 MW: AA28B116 CRC32:

Query Match 100.0%; Score 69; DB 8; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.93e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
 |||||  
 QY 1 RYRPGTVAL 9

RESULT 3  
 ID Q43031 PRELIMINARY; PRT: 56 AA.

AC Q43031;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE HISTONE H3 (FRAGMENT).  
 OS PSEUDOMINIA COMOSA.  
 OC EUKARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;  
 OC ROSIDAE; FABALES; FABACEAE; PAPILIONOIDEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDJ DATA BANKS.  
 CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
 CC -1- IN NUCLEOSOME FORMATION.  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 CC EMBL: U47408; G1208725;  
 DR EMBL: U47408; G1208725;  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.  
 FT NON\_TER 1 1  
 SO SEQUENCE 56 AA: 6522 MW: AB5A375A CRC32:

Query Match 100.0%; Score 69; DB 8; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.93e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
 |||||  
 QY 1 RYRPGTVAL 9

RESULT 4  
 ID Q42743 PRELIMINARY; PRT: 56 AA.

AC Q42743; Q42741; Q42746;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE HISTONE H3 (FRAGMENT).  
 OS GLYCINE AFF. TABACINA.  
 OC EUKARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;  
 OC ROSIDAE; FABALES; FABACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE: 97131650.  
 RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
 RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).  
 CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
 CC -1- IN NUCLEOSOME FORMATION.  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 CC EMBL: U47393; G1208678;  
 DR EMBL: U47375; G1208646;

DR EMBL: U47410; G1208717;  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.  
 FT NON\_TER 1 1  
 SO SEQUENCE 56 AA: 6506 MW: AA28B116 CRC32:

Query Match 100.0%; Score 69; DB 8; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.93e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
 |||||  
 QY 1 RYRPGTVAL 9

RESULT 5  
 ID Q43029 PRELIMINARY; PRT: 56 AA.

AC Q43029;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE HISTONE H3 (FRAGMENT).  
 OS PSEUDOMINIA COMOSA.  
 OC EUKARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;  
 OC ROSIDAE; FABALES; FABACEAE; PAPILIONOIDEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE: 97131650.  
 RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
 RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).  
 DR EMBL: U47373; G1208668;  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.  
 FT NON\_TER 1 1  
 SO SEQUENCE 56 AA: 6538 MW: 84A0869F CRC32:

Query Match 100.0%; Score 69; DB 8; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.93e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
 |||||  
 QY 1 RYRPGTVAL 9

RESULT 6  
 ID Q42724 PRELIMINARY; PRT: 56 AA.

AC Q42724;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE HISTONE H3 (FRAGMENT).  
 OS DIMASTIA VILLOSA.  
 OC EUKARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;  
 OC ROSIDAE; FABALES; FABACEAE; PAPILIONOIDEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE: 97131650.  
 RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
 RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).  
 DR EMBL: U47383; G1208644;  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.  
 FT NON\_TER 1 1  
 SO SEQUENCE 56 AA: 6453 MW: BID783C0 CRC32:

Query Match 100.0%; Score 69; DB 8; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.93e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
 |||||  
 QY 1 RYRPGTVAL 9

```

RESULT 7
ID 043030 PRELIMINARY: PRT: 56 AA.
AC 043030:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HISTONE H3 (FRAGMENT)
OS PSEUDOMINIA COMOSA.
OC EUKARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;
OC ROSIDAE; FABALES; FABACEAE; PAPILIONOIDEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97131650.
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;
RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).
DR EMBL: U47391; G1208713; -
KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 56 AA: 6506 MW; AA28B116 CRC32;

Query Match
Best Local Similarity 100.0%; Score 69; DB 8; Length 56;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
OY 1 RYRPGTVAL 9

RESULT 8
ID 042755 PRELIMINARY: PRT: 56 AA.
AC 042755: 042748;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HISTONE H3 (FRAGMENT)
OS GLYCINE CURVATA.
OC EUKARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;
OC ROSIDAE; FABALES; FABACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97131650.
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;
RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).
DR EMBL: U47360; G1213293; -
KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 56 AA: 6506 MW; AA28B116 CRC32;

Query Match
Best Local Similarity 100.0%; Score 69; DB 8; Length 56;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
OY 1 RYRPGTVAL 9

RESULT 9
ID 042821 PRELIMINARY: PRT: 56 AA.
AC 042821:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HISTONE H3 (FRAGMENT)
OS GLYCINE MAX (SOYBEAN).
OC EUKARYOTA; PLANTAE; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE; FABALES;
OC FABACEAE.

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97131650.
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;
RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).
DR EMBL: U47403; G1208707; -
KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 56 AA: 6490 MW; 59A39A7B CRC32;

Query Match
Best Local Similarity 100.0%; Score 69; DB 8; Length 56;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
OY 1 RYRPGTVAL 9

RESULT 10
ID 043228 PRELIMINARY: PRT: 56 AA.
AC 043228:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HISTONE H3 (FRAGMENT)
OS TERAMUS LABIALIS.
OC EUKARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;
OC ROSIDAE; FABALES; FABACEAE; PAPILIONOIDEAE.
RN [1]
RP SEQUENCE FROM N.A.
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
CC IN NUCLEOSOME FORMATION.
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
DR EMBL: U47411; G1208727; -
KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 56 AA: 6537 MW; 2E19A8E7 CRC32;

Query Match
Best Local Similarity 100.0%; Score 69; DB 8; Length 56;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
OY 1 RYRPGTVAL 9

RESULT 11
ID 042820 PRELIMINARY: PRT: 56 AA.
AC 042820: 042790; 042816; 042819;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HISTONE H3 (FRAGMENT)
OS GLYCINE MAX (SOYBEAN).
OC EUKARYOTA; PLANTAE; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE; FABALES;
OC FABACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97131650.
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;
RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).
DR EMBL: U47402; G1208705; -
DR EMBL: U47406; G1208703; -
DR EMBL: U47365; G1213311; -
DR EMBL: U47386; G1208660; -
KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.

```

PT NON\_TER 1  
 FT NON\_TER 56  
 SO SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 8; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.93e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
 |||||  
 QY 1 RYRPGTVAL 9

RESULT 12  
 ID 042759 PRELIMINARY; PRT; 56 AA.

AC 042759: 042757; DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
 RX MEDLINE; 97131650.  
 RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
 RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).  
 DR EMBL; U47401: G1208690; -;  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SO SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 8; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.93e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
 |||||  
 QY 1 RYRPGTVAL 9

RESULT 13  
 ID 042789 PRELIMINARY; PRT; 56 AA.

AC 042789: 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE HISTONE H3 (FRAGMENT).  
 OS GLYCINE MICROPHYLIA.  
 OC EUKARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;  
 OC ROSIDAE; FABALES; FABACEAE.  
 RN [1].  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97131650.  
 RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
 RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).  
 DR EMBL; U47390: G1208701; -;  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SO SEQUENCE 56 AA; 6534 MW; 91748156 CRC32;

Query Match 100.0%; Score 69; DB 8; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.93e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
 |||||  
 QY 1 RYRPGTVAL 9

RESULT 14  
 ID 039762 PRELIMINARY; PRT; 56 AA.

AC 039762: 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE HISTONE H3 (FRAGMENT).  
 OS GLYCINE ARENARIA.  
 OC EUKARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;  
 OC ROSIDAE; FABALES; FABACEAE.  
 RN [1].  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97131650.  
 RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
 RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).  
 DR EMBL; U47357: G1213287; -;  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SO SEQUENCE 56 AA; 6472 MW; 9077D730 CRC32;

Query Match 100.0%; Score 69; DB 8; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.93e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
 |||||  
 QY 1 RYRPGTVAL 9

RESULT 15  
 ID 042745 PRELIMINARY; PRT; 56 AA.

AC 042745: 042740: DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
 RX MEDLINE; 97131650.  
 RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
 RL MOL. PHYLOGENET. EVOL. 6:438-447(1996).  
 DR EMBL; U47397: G1208682; -;  
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING; NUCLEOSOME CORE.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SO SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 8; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.93e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
 |||||  
 QY 1 RYRPGTVAL 9

Search completed: Fri Sep 11 13:19:12 1998  
 Job time : 22 secs.



\*\*\*\*\*  
M P S R E H  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSPrch JP protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:19:29 1998; Maspar time 1.16 seconds

Tabular output not generated. 54,826 Million cell updates/sec

Title: >US-08-452-843-11  
Description: (1-9) from US08452843.pep  
Perfect Score: 69  
Sequence: 1 RYRGTVALL 9

Scoring table: PAM 150  
Gap 15

Searched: 77021 seqs, 7058996 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5.COMB 2:PCP9\_COMB 3:backfiles1

Statistics: Mean 16.005; Variance 41.780; scale 0.383

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No.              |
|------------|-------|-------------|--------|----|-------------|------------------------|
| 1          | 45    | 65.2        | 377    | 2  | PCT-US95-1  | Sequence 4, Applicatio |
| 2          | 45    | 65.2        | 824    | 1  | US-08-221-  | Sequence 3, Applicatio |
| 3          | 44    | 63.8        | 439    | 1  | US-08-553-  | Sequence 2, Applicatio |
| 4          | 44    | 63.8        | 655    | 1  | US-07-736-  | Sequence 2, Applicatio |
| 5          | 44    | 63.8        | 678    | 2  | PCT-US93-0  | Sequence 3, Applicatio |
| 6          | 44    | 63.8        | 2466   | 1  | US-07-915-  | Sequence 2, Applicatio |
| 7          | 43    | 62.3        | 211    | 1  | US-08-118-  | Sequence 19, Applicati |
| 8          | 43    | 62.3        | 322    | 1  | US-08-118-  | Sequence 75, Applicati |
| 9          | 43    | 62.3        | 322    | 2  | PCT-US93-0  | Patent No. 5474933.    |
| 10         | 43    | 62.3        | 359    | 3  | 5474933-8   | Sequence 2, Applicatio |
| 11         | 43    | 62.3        | 391    | 1  | US-08-417-  | Sequence 14, Applicati |
| 12         | 43    | 62.3        | 391    | 1  | US-08-417-  | Sequence 4, Applicatio |
| 13         | 43    | 62.3        | 391    | 1  | US-07-816-  | Sequence 2, Applicatio |
| 14         | 43    | 62.3        | 391    | 1  | US-07-816-  | Sequence 4, Applicatio |
| 15         | 43    | 62.3        | 596    | 1  | US-08-565-  | Sequence 11, Applicati |
| 16         | 43    | 62.3        | 826    | 1  | US-07-638-  | Sequence 2, Applicatio |
| 17         | 43    | 62.3        | 826    | 2  | PCT-US93-0  | Sequence 2, Applicatio |
| 18         | 43    | 62.3        | 826    | 2  | PCT-US93-0  | Sequence 2, Applicatio |
| 19         | 42    | 60.9        | 879    | 1  | US-08-554-  | Sequence 1, Applicatio |
| 20         | 42    | 60.9        | 1170   | 1  | US-08-032-  | Sequence 2, Applicatio |
| 21         | 42    | 60.9        | 1261   | 1  | US-08-764-  | Sequence 26, Applicati |
| 22         | 42    | 60.9        | 3218   | 1  | US-08-764-  | Sequence 27, Applicati |
| 23         | 41    | 59.4        | 551    | 1  | US-08-484-  | Sequence 15, Applicati |

| RESULT | ID | PCT-US95-10403-4 | STANDARD: | PRT: | 377 AA.    |
|--------|----|------------------|-----------|------|------------|
| 24     | 41 | 59.4             | 553       | 1    | US-08-565- |
| 25     | 41 | 59.4             | 718       | 2    | PCT-US95-0 |
| 26     | 41 | 59.4             | 963       | 1    | US-08-537- |
| 27     | 41 | 59.4             | 1141      | 1    | US-08-563- |
| 28     | 41 | 59.4             | 3038      | 1    | US-08-450- |
| 29     | 40 | 58.0             | 29        | 1    | US-08-340- |
| 30     | 40 | 58.0             | 235       | 1    | US-08-591- |
| 31     | 40 | 58.0             | 278       | 1    | US-08-441- |
| 32     | 40 | 58.0             | 278       | 1    | US-07-921- |
| 33     | 40 | 58.0             | 255       | 1    | US-08-118- |
| 34     | 40 | 58.0             | 295       | 2    | PCT-US93-0 |
| 35     | 40 | 58.0             | 323       | 1    | US-08-591- |
| 36     | 40 | 58.0             | 331       | 1    | US-08-319- |
| 37     | 40 | 58.0             | 375       | 1    | US-08-736- |
| 38     | 40 | 58.0             | 1147      | 1    | US-08-144- |
| 39     | 40 | 58.0             | 1165      | 1    | US-08-144- |
| 40     | 39 | 56.5             | 12        | 3    | 5512660-5  |
| 41     | 39 | 56.5             | 240       | 3    | 5472691-2  |
| 42     | 39 | 56.5             | 256       | 1    | US-07-906- |
| 43     | 39 | 56.5             | 274       | 3    | 548533-2   |
| 44     | 39 | 56.5             | 274       | 3    | 5512660-2  |
| 45     | 39 | 56.5             | 491       | 1    | US-08-462- |

## ALIGNMENTS

Sequence 4, Application PC/TUS9510403

Sequence 4, Application PC/TUS9510403

GENERAL INFORMATION:

APPLICANT: The General Hospital Corporation

TITLE OF INVENTION: GA4 DNA, Protein and Methods of Use

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sterne, Kessler, Goldstein and Fox

STREET: 1100 New York Avenue, Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: U.S.

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10403

FILING DATE: 15 August 1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/291,939

FILING DATE: 16 August 1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Cimbal, Michele A.

REGISTRATION NUMBER: 33,851

REFERENCE/DOCKET NUMBER: 0609, 408PC00

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 377 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

| RESULT    | 3                |
|-----------|------------------|
| ID        | US-08-553-999B-2 |
| STANDARD; | PRT;             |
|           | 439 AA           |

CC ADDRESSEE: E11 Lilly And Company

CC ADDRESSEE: Eli Lilly And Company  
CC STREET: Lilly corporate center  
CC

```
CC NAME: DOROTHY R. AUTH
CC REGISTRATION NUMBER: P-36,434
CC REFERENCE/DOCKET NUMBER: 2026-4010 PCT
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-758-4800
CC TELEFAX: 212-751-6849
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 678
CC TYPE: AMINO ACID
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: No
CC ORIGINAL SOURCE:
CC ORGANISM: Drosophila melanogaster
CC STRAIN: Oregon R
CC INDIVIDUAL ISOLATE:
CC DEVELOPMENTAL STAGE: embryo
CC HAPLOTYPE:
CC TISSUE TYPE:
CC CELL LINE:
CC ORGANELLE:
CC FEATURE:
CC NAME/KEY: Dorsal protein
CC LOCATION:
CC IDENTIFICATION METHOD:
CC OTHER INFORMATION: D.melanogaster
CC OTHER INFORMATION: embryonic polarity (dorsal) protein
CC OTHER INFORMATION: containing region of high similarity
CC OTHER INFORMATION: with proteins of Rel family.
CC PUBLICATION INFORMATION:
CC AUTHORS: Steward, R.
CC TITLE: Dorsal, an embryonic polarity
CC TITLE: gene in Drosophila, is homologous to
CC TITLE: the vertebrate proto-oncogene, c-rel.
CC JOURNAL: Science
CC VOLUME: 238
CC ISSUE:
CC PAGES: 692-694
CC DATE: 1987
CC DOCUMENT NUMBER:
CC FILING DATE:
CC PUBLICATION DATE:
CC RELEVANT RESIDUES IN SEQ ID NO:
SQ SEQUENCE 678 AA; 75502 MM; 2127035 CN;

Query Match 63.8%; Score 44; DB 2; Length 678;
Best Local Similarity 44.4%; Pred. NO. 6.69e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 177 RFPSTIDL 185
QY 1 RYRGTVAL 9

RESULT 6
ID PCT-US94-09943-2 STANDARD; PRT; 2466 AA.
XX AC xxxxxx
XX DT
XX XX
XX XX
XX XX
Sequence 2, Application PC/TUS9409943
DE Sequence 2, Application PC/TUS9409943
CC GENERAL INFORMATION:
CC APPLICANT:
CC APPLICANT:
CC APPLICANT:
CC APPLICANT:
CC APPLICANT:
```

```

CC      OPERATING SYSTEM:  PC-DOS/MS-DOS
CC      SOFTWARE:  Patentin Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER:  US/07/915,966C
CC      FILING DATE:  17-JUL-1992
CC      CLASSIFICATION:  435
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME:  Matthews, Gale M.
CC      REGISTRATION NUMBER:  32,269
CC      REFERENCE/DOCKET NUMBER:  31,829-00
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE:  201-683-2134
CC      TELEFAX:  201-683-4117
CC      INFORMATION FOR SEO ID NO:  19:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH:  211 amino acids
CC      TYPE:  amino acid
CC      STRANDEDNESS:  single
CC      TOPOLOGY:  linear
CC      MOLECULE TYPE:  peptide
CC      HYPOTHEITICAL:  NO
CC      ANTI-SENSE:  NO
CC      FRAGMENT TYPE:  internal
CC      ORIGINAL SOURCE:
CC      ORGANISM:  Human
CC      SEQUENCE  211 AA; 23793 MW; 259278 CN;

Query Match          62.3%;  SCORE 43;  DB 1;  Length 211;
Best Local Similarity 75.0%;  Pred. No. 8.79e+01;
Matches  6;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Db      63  RYRPTVA 70
      ||| |||
Oy      1  RYRGTVA 8

RESULT  8
ID      US-08-118-270-75      STANDARD;      PRT;  322 AA.
XX      xxxxxx

Sequence 75, Application US/08118270
CC      Patent No. 5508384
CC      GENERAL INFORMATION:
CC      APPLICANT:  Murphy, Randall B.
CC      APPLICANT:  Schuster, David I.
CC      TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
CC      TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
CC      NUMBER OF SEQUENCES:  348
CC      CORRESPONDENCE ADDRESSES:
CC      ADDRESSEE:  BROWDY AND NEIMARK
CC      STREET:  419 Seventh Street, N.W., Suite 300
CC      CITY:  Washington
CC      STATE:  D.C.
CC      COUNTRY:  USA
CC      ZIP:  20004
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE:  Floppy disk
CC      COMPUTER:  IBM PC compatible
CC      OPERATING SYSTEM:  PC-DOS/MS-DOS
CC      SOFTWARE:  Patentin Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER:  US/08/118,270
CC      FILING DATE:  09-SEP-1993
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 07/943,236
CC      FILING DATE:  10-SEP-1992
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME:  Townsend, Kevin G.

```

```

CC      REGISTRATION NUMBER: 34.033
CC
CC      REFERENCE/DOCKET NUMBER: MORPH-2
CC
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: 202-628-5197
CC      TELEFAX: 202-737-3528
CC
CC      TELEX: 248633
CC
CC      INFORMATION FOR SEQ ID NO: 75:
CC
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 322 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
CC
SO      SEQUENCE 322 AA: 36296 MW: 560567 CN:

```

|    |     |          |     |
|----|-----|----------|-----|
| Db | 106 | RYRRPTVA | 113 |
|    |     |          |     |
| QY | 1   | RYRPGTVA | 8   |

AC XXXXXXXX  
YV

DE . Sequence 75, Application PC/TUS9308528

CC Sequence 75, Application PC/TUS9308528  
CC GENERAL INFORMATION:  
CC APPLICANT: New York University  
CC TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
CC TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
CC NUMBER OF SEQUENCES: 348  
CC CORRESPONDENCE ADDRESS:  
CC

CC ADDRESS: BROWDY AND NEIMARK  
CC STREET: 419 Seventh Street, N.W., Suite 300  
CC City: Washington  
CC STATE: D.C.  
CC COUNTRY: USA

```
CC ; COMPUTER READABLE FORM:
CC
CC MEDIUM TYPE: Floppy disk
CC
CC COMPUTER: IBM PC compatible
CC
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC
CC CURRENT APPLICATION DATA:
```

CC APPLICATION NUMBER: PCT/US93/08528  
CC FILING DATE: 09-SEP-1993  
CC

CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/943,236  
CC

CC FILING DATE: 10-SEP-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Edward J. Davis

CC NAME: IOWSEND, KEVIN G.  
CC REGISTRATION NUMBER: 34,033  
CC REFERENCE/DOCKET NUMBER: MURPHY-3 BCT

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528  
TELEX: 248633

CC INFORMATION FOR SEQ ID NO: 75:  
CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 322 amino acids  
CC TYPE: amino acid  
CC

MOLECULE TYPE: peptide

SEQUENCE 322 AA; 36296 MW; 560567 CN;

|    |     |          |     |
|----|-----|----------|-----|
| Db | 106 | RYRPTVA  | 113 |
|    |     |          |     |
| QY | 1   | RYRPGTVA | 8   |

|        |           |           |             |
|--------|-----------|-----------|-------------|
| RESULT | 10        |           |             |
| ID     | 5474933-8 | STANDARD; | PRT; 389 AA |

XX  
DT  
XX

01-JAN-1900

Patent No. 5474933.

Patent No. 5474933  
APPLICANT: FIN

CC TITLE OF INVENTION: MARINE MELA GENE  
CC  
CC NUMBER OF SEQUENCES: 9  
CC

CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/148,945  
CC

CC FILING DATE: 08-NOV-1993  
CC PRIOR APPLICATION DATA:  
CC

CC APPLICATION NUMBER: 974,837  
CC FILING DATE: 10-NOV-1992  
CC ADDITIONAL WRAPD: 405 0044

CC APPLICATION NUMBER: 496,804  
CC FILING DATE: 21-MAR-1990  
CC SEQ. ID NO.: 8

```

CC      SEQ ID NO: 1
CC      LENGTH: 359
SO      SEQUENCE 389 AA: 43629 MW: 831091 CN:

```

Query Match      62.38;    Score 43;    DB 3;    length 359;

Best Local Similarity 50.0%; Pred. NO. 8.79e+01;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 49 YRQGAJNL 56

|    |   |          |   |
|----|---|----------|---|
| QY | 2 | YRPGTVAL | 9 |
|----|---|----------|---|

## RESULT 11

ID US-08-417-103-2  
XX

AC                    xxxxxxxx  
XX                    xx

DT  
XX

|    |                                     |
|----|-------------------------------------|
| DE | sequence 2, Application US/08417103 |
| CC | sequence 3 Application US/08417103  |
| XX |                                     |

sequence 2, Appl. No. 5723299

CC APPLICANT: Bell, Graeme I.  
CC  
CC APPLICANT: Yamada, Yutshiro  
CC

CC APPLICANT: Selno, Susumu  
CC TITLE OF INVENTION: SOMATOSTATIN RECEPTORS  
CC

CC NUMBER OF SEQUENCES: 16  
CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Arnold, White & Durkee  
CC STREET: P.O. Box 4433  
CC

CITY: HOUSTON  
STATE: Texas  
COUNTY: HARRIS

CC COMM: United States of America  
CC ZIP: 77210  
CC COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC

CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/417,103  
CC FILING DATE: 05-APR-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/816,283  
CC FILING DATE: 01-DEC-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Wilson, Mark B.  
CC REGISTRATION NUMBER: 37,259  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (512) 418-3000  
CC TELEFAX: (512) 474-7577  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 391 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 391 AA; 42686 MW; 839883 CN;  
SO  
Query Match 62.3%; Score 43; DB 1; Length 391;  
Best Local Similarity 75.0%; Pred. No. 8.79e+01;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
DB 167 RYRPTVA 174  
QY 1 RYRPTVA 8  
RESULT 12  
ID US-08-417-103-14 STANDARD; PRT; 391 AA.  
XX xxxxxx  
XX  
XX  
DE Sequence 14, Application US/08417103  
CC Sequence 14, Application US/08417103  
CC Patent No. 5723299  
CC GENERAL INFORMATION:  
CC APPLICANT: Bell, Graeme I.  
CC APPLICANT: Yamada, Yuichiro  
CC APPLICANT: Saino, Susumu  
CC TITLE OF INVENTION: SOMATOSTATIN RECEPTORS  
CC NUMBER OF SEQUENCES: 16  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/417,103  
CC FILING DATE: 05-APR-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/816,283  
CC FILING DATE: 01-DEC-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Wilson, Mark B.  
CC REGISTRATION NUMBER: 37,259  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (512) 418-3000  
CC TELEFAX: (512) 474-7577  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 391 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 391 AA; 42686 MW; 839883 CN;  
SO  
Query Match 62.3%; Score 43; DB 1; Length 391;  
Best Local Similarity 75.0%; Pred. No. 8.79e+01;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CC TELEFAX: (512) 474-7577  
CC INFORMATION FOR SEQ ID NO: 14:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 391 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 391 AA; 42686 MW; 839883 CN;  
SO  
Query Match 62.3%; Score 43; DB 1; Length 391;  
Best Local Similarity 75.0%; Pred. No. 8.79e+01;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
DB 167 RYRPTVA 174  
QY 1 RYRPTVA 8  
RESULT 13  
ID US-08-417-103-4 STANDARD; PRT; 391 AA.  
XX xxxxxx  
XX  
XX  
DE Sequence 4, Application US/08417103  
CC Sequence 4, Application US/08417103  
CC Patent No. 5723299  
CC GENERAL INFORMATION:  
CC APPLICANT: Bell, Graeme I.  
CC APPLICANT: Yamada, Yuichiro  
CC APPLICANT: Saino, Susumu  
CC TITLE OF INVENTION: SOMATOSTATIN RECEPTORS  
CC NUMBER OF SEQUENCES: 16  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/417,103  
CC FILING DATE: 05-APR-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/816,283  
CC FILING DATE: 01-DEC-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Wilson, Mark B.  
CC REGISTRATION NUMBER: 37,259  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (512) 418-3000  
CC TELEFAX: (512) 474-7577  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 391 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 391 AA; 42718 MW; 833561 CN;  
SO  
Query Match 62.3%; Score 43; DB 1; Length 391;  
Best Local Similarity 75.0%; Pred. No. 8.79e+01;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 167 RYRPTVA 174  
111111  
QY 1 RYRPTVA 8

RESULT 14  
ID US-07-816-283-2 STANDARD; PRT; 391 AA.  
XX  
AC xxxxxx  
XX  
DT

Sequence 2, Application US/07816283  
XX  
CC Sequence 2, Application US/07816283  
CC Patent No. 5436155  
CC GENERAL INFORMATION:  
CC APPLICANT: Bell, Graeme I.  
CC APPLICANT: Yamada, Yuichiro  
CC APPLICANT: Selnou, Susumu  
CC TITLE OF INVENTION: SOMATOSTATIN RECEPTORS  
CC NUMBER OF SEQUENCES: 12  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Arnold, White & Durkee  
CC STREET: PO Box 4433  
CC CITY: Houston  
CC STATE: Texas  
CC COUNTRY: USA  
CC ZIP: 77210  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/816,283  
CC FILING DATE: 19911231  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Mcdaniel, C. Steven  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 713-787-1400  
CC TELEFAX: 713-789-2679  
CC TELEX: 79-0924  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 391 amino acids  
CC TYPE: AMINO ACID  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 391 AA: 42686 MW; 839883 CN;

Query Match 62.3%; Score 43; DB 1; Length 391;  
Best Local Similarity 75.0%; Pred. No. 8.79e+01;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 167 RYRPTVA 174  
111111  
QY 1 RYRPTVA 8

RESULT 15  
ID US-07-816-283-4 STANDARD; PRT; 391 AA.  
XX  
AC xxxxxx  
XX  
DT

Sequence 4, Application US/07816283  
XX  
CC Sequence 4, Application US/07816283  
CC Patent No. 5436155  
CC GENERAL INFORMATION:

CC APPLICANT: Bell, Graeme I.  
CC APPLICANT: Yamada, Yuichiro  
CC APPLICANT: Selnou, Susumu  
CC TITLE OF INVENTION: SOMATOSTATIN RECEPTORS  
CC NUMBER OF SEQUENCES: 12  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Arnold, White & Durkee  
CC STREET: PO Box 4433  
CC CITY: Houston  
CC STATE: Texas  
CC COUNTRY: USA  
CC ZIP: 77210  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/816,283  
CC FILING DATE: 19911231  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Mcdaniel, C. Steven  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 713-787-1400  
CC TELEFAX: 713-789-2679  
CC TELEX: 79-0924  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 391 amino acids  
CC TYPE: AMINO ACID  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 391 AA: 42718 MW; 833561 CN;

Query Match 62.3%; Score 43; DB 1; Length 391;  
Best Local Similarity 75.0%; Pred. No. 8.79e+01;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 167 RYRPTVA 174  
111111  
QY 1 RYRPTVA 8

Search completed: Fri Sep 11 13:19:35 1998  
Job time : 6 secs.

**THIS PAGE BLANK (USPTO)**





CC FEATURE:  
CC NAME/KEY: Peptide  
CC LOCATION: 1..9  
CC OTHER INFORMATION: /note="Histone H3.3"  
SO SEQUENCE 9 AA: 1032 MW: 495 CN:

Query Match 100.0%; Score 69; DB 8; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.52e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 RYRPGTVAL 9  
QY 1 RYRPGTVAL 9

RESULT 2  
ID US-08-424-484-77 STANDARD; PRT; 9 AA.

AC xxxxxx

DE Sequence 77, Application US/08424484

CC Sequence 77, Application US/08424484

CC GENERAL INFORMATION:

CC APPLICANT: R tzsckhe, Olaf

CC APPLICANT: Falk, Kirsten

CC APPLICANT: Stevanovi, Stefan

CC APPLICANT: Ramnensee, Hans-Georg

CC APPLICANT: Jung, G nther

CC TITLE OF INVENTION: DETERMINATION OF PEPTIDE MOTIFS ON MHC

CC NUMBER OF SEQUENCES: 117

CC CORRESPONDENCE ADDRESSES:

CC ADDRESSEE: Nikaido, Marmelstein, Murray & Oram

CC STREET: 655 Fifteenth Street N.W. Suite 330

CC CITY: Washington

CC STATE: D.C.

CC COUNTRY: U.S.A.

CC ZIP: 20005-5701

CC COMPUTER, READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC OPERATING SYSTEM: IBM PC compatible

CC SOFTWARE: Patentin Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/424,484

CC FILING DATE: 1-SEP-1995

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Kites, Monica C.

CC REGISTRATION NUMBER: 36,105

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (202)638-5000

CC TELEFAX: (202)638-4810

CC INFORMATION FOR SEQ ID NO: 77:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 9 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: Peptide

CC SEQUENCE 9 AA: 1032 MW: 495 CN:

Query Match 100.0%; Score 69; DB 8; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.52e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 RYRPGTVAL 9  
QY 1 RYRPGTVAL 9

RESULT 3  
ID US-08-390-207-12 STANDARD; PRT; 70 AA.  
AC xxxxxx

DE Sequence 12, Application US/08390207

CC Sequence 12, Application US/08390207

CC GENERAL INFORMATION:

CC APPLICANT: KATO, Seishi

CC APPLICANT: OH, Suwan

CC APPLICANT: SEKINE, Shingo

CC APPLICANT: KIM, Namsoon

CC APPLICANT: KATO, Takeo

CC APPLICANT: IMAHORI, Akiyo

CC TITLE OF INVENTION: HUMAN CDNA AND PROTEINS ENCODED THEREBY

CC NUMBER OF SEQUENCES: 253

CC CORRESPONDENCE ADDRESSES:

CC ADDRESSEE: Foley & Lardner

CC STREET: 3000 K Street, N.W., Suite 500

CC CITY: Washington

CC STATE: D.C.

CC COUNTRY: USA

CC ZIP: 20007-5109

CC COMPUTER, READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC OPERATING SYSTEM: IBM PC compatible

CC SOFTWARE: Patentin Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/390,207

CC FILING DATE: 16-FEB-1995

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: JP 4-208077

CC FILING DATE: 04-AUG-1992

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: JP 4-327619

CC FILING DATE: 13-NOV-1992

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: JP 5-61431

CC FILING DATE: 26-FEB-1993

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: WO PCT/JP93/01095

CC FILING DATE: 04-AUG-1993

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/379,441

CC FILING DATE: 03-FEB-1995

CC ATTORNEY/AGENT INFORMATION:

CC NAME: WEGNER, Harold C.

CC REGISTRATION NUMBER: 25,258

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (202)672-5300

CC TELEFAX: (202)672-5399

CC TELEEX: 904136

CC INFORMATION FOR SEQ ID NO: 12:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 70 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 70 AA: 7916 MW: 24433 CN:

Query Match 100.0%; Score 69; DB 7; Length 70;  
Best Local Similarity 100.0%; Pred. No. 1.52e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 41 RYRPGTVAL 49  
QY 1 RYRPGTVAL 9

US-08-452-843-11.rap

```

XX Sequence 12, Application US/08716249
DE
XX
CC Sequence 12, Application US/08716249
CC GENERAL INFORMATION:
CC APPLICANT: Guichard, Gilles
CC APPLICANT: Muller, Sylviane
CC APPLICANT: Briand, Jean-Paul
CC APPLICANT: Regemortel, Marc
CC TITLE OF INVENTION: Retropeptides, Antibodies Thereto, and
CC TITLE OF INVENTION: Uses Thereof for Vaccination and In Vitro Diagnosis
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Spencer & Frank
CC STREET: 1100 New York Avenue, Suite 300E
CC CITY: Washington, D.C.
CC COUNTRY: USA
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/716,249
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: WO PCT/FR95/00292
CC FILING DATE: 13-MAR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Calvetti, Frederick F.
CC REGISTRATION NUMBER: 28,557
CC REFERENCE/DOCKET NUMBER: GROFO 7001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202)414-4000
CC TELEFAX: (202)414-4040
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 18 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHEICAL: NO
CC ANTI-SENSE: NO
CC SQ SEQUENCE 18 AA: 1880 MW; 1921 CN;

Query Match 72.5%; Score 50; DB 11; Length 18;
Best Local Similarity 100.0%; Pred. No. 5.40e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 13 RYRPGT 18
OY 1 RYRPGT 6

RESULT 6 STANDARD: PRT: 566 AA.
ID US-08-886-333-7
XX
XX xxxxxx
XX
XX
DE Sequence 7, Application US/08886333
CC
CC GENERAL INFORMATION:
CC APPLICANT: Collige, Alain
CC APPLICANT: Lapiere, Charles M.
CC APPLICANT: Prockop, Darwin J.
CC TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
CC TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
CC NUMBER OF SEQUENCES: 17
CC

```

```

CC      FILING DATE: 03-MAY-1996
CC      CLASSIFICATION: 435
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: PCT/GB95/02875
CC      FILING DATE: 11-DEC-1995
CC      CLASSIFICATION: 435
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Pabst, Patrea L.
CC      REGISTRATION NUMBER: 31,284
CC      REFERENCE/DOCKET NUMBER: RPMS 101
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (404) 873-8794
CC      TELEFAX: (404) 873-8795
CC      INFORMATION FOR SEQ ID NO: 215:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 41 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      HYPOTHETICAL: NO
CC      SEQUENCE 41 AA: 4207 MW: 9889 CN:
CC
CC      Query Match          66.7%  Score 46;  DB 10;  Length 41;
CC      Best Local Similarity 62.5%;  Pred. No. 1.71e+02;
CC      Matches          5;  Conservative          2;  Mismatches          1;  Indels          0;  Gaps          0;
CC
CC      Db
CC      1 YRTGSVGL 8
CC      |||:|:|
CC      QY          2 YRPTVAL 9
CC
CC      RESULT      8
CC      ID          US-07-661-071A-2          STANDARD;          PRT;          45 AA.
CC      XX          xxxxxx
CC      DT
CC      DT
CC      Sequence 2, Application US/07661071A
CC      DE
CC      CC      Sequence 2, Application US/07661071A
CC      CC      GENERAL INFORMATION:
CC      CC      APPLICANT: Zamecnik, Paul C
CC      CC      APPLICANT: Agrawal, Sudhir
CC      CC      TITLE OF INVENTION: Diagnosing Cystic Fibrosis and other
CC      CC      TITLE OF INVENTION: Genetic Diseases Using Fluorescence Resonance Energy
CC      CC      TITLE OF INVENTION:
CC      CC      NUMBER OF SEQUENCES: 4
CC      CC      CORRESPONDENCE ADDRESS:
CC      CC      ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
CC      CC      STREET: 2 Millitia Drive
CC      CC      City: Lexington
CC      CC      STATE: MA
CC      CC      COUNTRY: USA
CC      CC      ZIP: 02173
CC      CC      COMPUTER READABLE FORM:
CC      CC      MEDIUM TYPE: Floppy disk
CC      CC      COMPUTER: IBM PC compatible
CC      CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      CC      SOFTWARE: PatentIn Release #1.0, Version #1.25
CC      CC      CURRENT APPLICATION DATA:
CC      CC      APPLICATION NUMBER: US/07/661,071A
CC      CC      FILING DATE: 19910226
CC      CC      CLASSIFICATION: 435
CC      CC      ATTORNEY/AGENT INFORMATION:
CC      CC      NAME: Granahan, Patricia
CC      CC      REGISTRATION NUMBER: 32,227
CC      CC      REFERENCE/DOCKET NUMBER: WFE9-02
CC      CC      TELECOMMUNICATION INFORMATION:
CC      CC      TELEPHONE: 617-861-6240
CC      CC      TELEFAX: 617-861-9540
CC      CC      INFORMATION FOR SEQ ID NO: 2:

```





CC NUMBER OF SEQUENCES: 539  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: ABBOTT LABORATORIES  
 CC STREET: ONE HUNDRED ABBOTT PARK ROAD  
 CC CITY: ABBOTT PARK  
 CC STATE: IL  
 CC COUNTRY: USA  
 CC ZIP: 60064-3500  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent in Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/344,185C  
 CC FILING DATE:  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: FOREMSKI, PRISCILLA E.  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 708-937-6365  
 CC TELEFAX: 708-938-2623  
 CC INFORMATION FOR SEQ ID NO: 94:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 112 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 112 AA: 12702 MW: 71745 CN;

```

Query Match          65.2%   Score 45; DB 7; Length 112;
Best Local Similarity 83.3%; Pred No. 2,28e+02;
Matches             5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Dd      73 YRPGSV 78      .
OY.     2 YRPGTV 7
1
RESULT 14
ID : US-08-424-550B-222 STANDARD; PRT; 112 AA.
XX AC xxxxxx
XX DF
XX DE Sequence 222, Application US/08424550B
CC CC
CC CC Sequence 222, Application US/08424550B
CC GENERAL INFORMATION:
CC APPLICANT: JOHN N. SIMONS
CC APPLICANT: TAMI J. PILOT-MATIAS
CC APPLICANT: GEORGE J. DAWSON
CC APPLICANT: GEORGE G. SCHLADDER
CC APPLICANT: SURESH M. DESAI
CC APPLICANT: THOMAS P. LEARY
CC APPLICANT: ANTHONY SCOTT MERRHOF
CC APPLICANT: JAMES C. ERKER
CC APPLICANT: SHERI L. BUIK
CC APPLICANT: ISA K. MOSHAMMAR
CC TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
CC TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
CC NUMBER OF SEQUENCES: 716
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
CC STREET: 100 ABBOTT PARK ROAD
CC CITY: ABBOTT PARK
CC STATE: IL
CC CC
CC CC COUNTRY: USA
CC CC ZIP: 60064-3500
CC CC COMPUTER READABLE FORM:
CC CC MEDIUM TYPE: Floppy disk
CC CC COMPUTER: IBM PC compatible

```

```

CC      OPERATING SYSTEM:  PC-DOS/MS-DOS
CC      SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER:  US/08/424,550B
CC      FILING DATE:
CC      CLASSIFICATION:  435435
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME:  POREMBSKI, PRISCILLA E.
CC      REGISTRATION NUMBER:  33,207
CC      REFERENCE/DOCKET NUMBER:  5527.PC.01
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE:  708-937-6365
CC      TELEFAX:  708-938-2623
CC      INFORMATION FOR SEQ ID NO:  222:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH:  112 amino acids
CC      TYPE:  amino acid
CC      TOPOLOGY:  linear
CC      MOLECULE TYPE:  protein
CC      SEQUENCE  112 AA;  12691 MW;  71649 CN;

```

```

Query Match Similarity      65.2%; Score 45; DB 8; Length 112;
Best Local Similarity      83.3%; Pred. No. 2.28e+02;
Matches       5; Conservative    1; Mismatches   0; Indels     0; Gaps     0;

Db          73 YRPGSV 78
            ||||:|
OY          2 YRPGTV 7

RESULT      15
ID           US-08-292-309-2        STANDARD; PRT; 237 AA.
XX
AC          xxxxxx
DT
DX
Sequence 2, Application US/08292309
DE
CC          Sequence 2, Application US/08292309
CC          GENERAL INFORMATION:
CC          APPLICANT: Lin, Lih-Ling
CC          TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
CC          TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
CC          NUMBER OF SEQUENCES: 14
CC          CORRESPONDENCE ADDRESS:
CC          ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
CC          STREET: 87 Cambridgepark Drive
CC          CITY: Cambridge
CC          STATE: MA
CC          COUNTRY: USA
CC          ZIP: 02140
CC          COMPUTER READABLE FORM:
CC          MEDIUM TYPE: Floppy disk
CC          COMPUTER: IBM PC compatible
CC          OPERATING SYSTEM: PC-DOS/MS-DOS
CC          SOFTWARE: PatentIn Release #1.0, Version #1.25
CC          CURRENT APPLICATION DATA:
CC          APPLICATION NUMBER: US/08/292,309
CC          FILING DATE:
CC          CLASSIFICATION: 424
CC          ATTORNEY/AGENT INFORMATION:
CC          NAME: Brown, Scott A.
CC          REGISTRATION NUMBER: 32,724
CC          REFERENCE/DOCKET NUMBER: GI 5231
CC          TELECOMMUNICATION INFORMATION:
CC          TELEPHONE: (617) 498-8224
CC          TELEFAX: (617) 876-5851
CC          INFORMATION FOR SEQ ID NO: 2:
CC          SEQUENCE CHARACTERISTICS:
CC          LENGTH: 237 amino acids

```

CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 237 AA; 26069 MW; 264328 CN;

Query Match 65.2%; Score 45; DB 6; Length 237;  
Best Local Similarity 100.0%; Pred. No. 2.28e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 7 RYRPG 11  
|||||

QY 1 RYRPG 5

Search completed: Fri Sep 11 13:20:35 1998  
Job time : 42 secs.



\*\*\*\*\*  
 W085R04H (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:13:24 1998; Maspar time 2.58 Seconds

Tabular output not generated. 56.425 Million cell updates/sec

Title: >US-08-452-843-10  
 Description: (1-9) from US08452843.pep  
 Perfect Score: 77  
 Sequence: 1 IPFPIVRYL 9

Scoring table: PAM 150  
 Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a.geneseq32  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29

Statistics: Mean 17.763; Variance 54.135; scale 0.328

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description | Pred. No.                      |
|------------|-------|-------------|-----------|----|-------------|--------------------------------|
| 1          | 77    | 100.0       | 9         | 18 | R89371      | Cw6 consensus peptide 1.13e+01 |
| 2          | 57    | 74.0        | 9         | 18 | R89370      | Cw6 consensus peptide 2.02e+01 |
| 3          | 56    | 72.7        | 15        | 26 | W38827      | Peptide resembling an          |
| 4          | 56    | 72.7        | 516       | 18 | W00183      | Cytochrome P450A2.             |
| 5          | 56    | 72.7        | 516       | 18 | R93167      | Human cytochrome P450          |
| 6          | 56    | 72.7        | 516       | 13 | R72360      | Human cytochrome P450          |
| 7          | 56    | 72.7        | 713       | 11 | R60101      | Canine zona pellucida          |
| 8          | 56    | 72.7        | 715       | 10 | R55198      | Canine zona pellucida          |
| 9          | 56    | 72.7        | 716       | 12 | R50532      | Canine zona pellucida          |
| 10         | 56    | 72.7        | 716       | 10 | R55200      | Feline zona pellucida          |
| 11         | 53    | 68.8        | 9         | 18 | R89369      | Feline zona pellucida          |
| 12         | 53    | 68.8        | 713       | 2  | R06998      | Cw6 consensus peptide          |
| 13         | 51    | 66.2        | 271       | 23 | W19612      | Mouse zeta protein exh         |
| 14         | 51    | 66.2        | 271       | 23 | W19219      | Human growth hormone           |
| 15         | 51    | 66.2        | 341       | 17 | R87029      | Human growth hormone           |
| 16         | 51    | 66.2        | 341       | 13 | R74391      | Varicella zoster virus         |
| 17         | 51    | 66.2        | 341       | 13 | R74389      | vzv thymidine-kinase.          |
| 18         | 51    | 66.2        | 341       | 13 | R74390      | vzv thymidine-kinase.          |

| ID | Score | Query Match | Length DB | ID | Description | Pred. No.               |
|----|-------|-------------|-----------|----|-------------|-------------------------|
| 19 | 51    | 66.2        | 353       | 23 | W19608      | Pig growth hormone se   |
| 20 | 51    | 66.2        | 353       | 23 | W19215      | Swine growth hormone    |
| 21 | 51    | 66.2        | 361       | 23 | W19217      | Human growth hormone    |
| 22 | 51    | 66.2        | 362       | 23 | W19610      | Human growth hormone    |
| 23 | 51    | 66.2        | 364       | 23 | W19220      | Human growth hormone se |
| 24 | 51    | 66.2        | 364       | 23 | W19613      | Rat growth hormone se   |
| 25 | 51    | 66.2        | 380       | 7  | R37595      | Sequence of microsome   |
| 26 | 51    | 66.2        | 463       | 8  | R41877      | Rat glucagon-like pep   |
| 27 | 51    | 66.2        | 713       | 10 | R55194      | Porcine zona pellucida  |
| 28 | 50    | 64.9        | 300       | 15 | R76772      | Film protein derived    |
| 29 | 50    | 64.9        | 398       | 14 | R74206      | Human death associate   |
| 30 | 49    | 63.6        | 463       | 13 | R70006      | Human glucagon-like 1   |
| 31 | 48    | 62.3        | 125       | 7  | R38224      | Sequence of polypepti   |
| 32 | 48    | 62.3        | 210       | 3  | R13499      | P. denitrificans COB H  |
| 33 | 48    | 62.3        | 352       | 4  | R20794      | EHV-4 TK protein.       |
| 34 | 48    | 62.3        | 484       | 2  | P71081      | Sequence encoded by V   |
| 35 | 48    | 62.3        | 742       | 14 | R74094      | Human zona pellucida    |
| 36 | 48    | 62.3        | 745       | 10 | R55206      | Human zona pellucida    |
| 37 | 48    | 62.3        | 2368      | 26 | W26663      | Yeast checkpoint cont   |
| 38 | 47    | 61.0        | 357       | 18 | W00497      | Papillomavirus E2 bin   |
| 39 | 47    | 61.0        | 482       | 21 | W14439      | Protein involved in c   |
| 40 | 47    | 61.0        | 491       | 2  | R27787      | Adrenodoxin reductase   |
| 41 | 47    | 61.0        | 497       | 12 | R66693      | Human adrenodoxin-red   |
| 42 | 47    | 61.0        | 497       | 2  | R27786      | Adrenodoxin reductase   |
| 43 | 47    | 61.0        | 1159      | 23 | W01570      | Protein encoded by CH   |
| 44 | 47    | 61.0        | 1164      | 23 | W01571      | Protein encoded by fu   |
| 45 | 46    | 59.7        | 512       | 18 | W00652      | Cytochrome P450A1.      |

## ALIGNMENTS

RESULT 1  
 ID R89371 standard; peptide: 9 AA.  
 AC R89371.  
 DE 18-SEP-1996 (first entry)  
 DE Cw6 consensus peptide derived immunogenic peptide #3.  
 KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
 KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
 KW hepatitis C.  
 OS Synthetic.  
 PN W09603140-A1.  
 PD 08-FEB-1996.  
 PE 21-JUL-1995; U09234.  
 PR 23-NOV-1994; US-344824.  
 PR 23-NOV-1994; US-344824.  
 PR 30-MAY-1995; US-452843.  
 PA (CYTE-) CYTEL CORP.  
 PI Sette A, Sidney J;  
 DR WPI: 96-116784/12.

PT Compsn. comprising immunogenic peptide with supermotif allowing more  
 than one HLA mol. to bind - used to induce CTL response in patient  
 PT and for in vivo and ex vivo therapeutic and diagnostic applications  
 PS Claim 2; Page 26; 32pp; English.  
 CC The sequences given in R89362-82 are immunogenic peptides which were  
 CC use in the composition of the invention. The composition comprises  
 CC an immunogenic peptide of 9-10 residues with a supermotif which  
 CC allows binding of more than one HLA molecule. It pref. comprises  
 CC two conserved residues, a first at the 2nd position from the N-  
 CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
 CC are used to induce a CTL response in a patient. They are also  
 CC useful in compositions for in vivo and ex vivo therapeutic and  
 CC diagnostic applications, e.g. the treatment of cancer and viral  
 CC infections, e.g. hepatitis B and C.  
 SQ Sequence 9 AA;

Query Match 100.0%; Score 77; DB 18; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.13e+01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 ipfpiyrl 9  
 1 ipfpiyrl 9  
 1 ipfpiyrl 9

RESULT 2

ID R89370 standard; peptide; 9 AA.

AC R89370;

DT 18-SEP-1996 (first entry)

DE C66 consensus peptide derived immunogenic peptide #2.

KW Immunogenic peptide; supermotif; HLA molecule; CTL response;

KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;

KW hepatitis C.

OS Synthetic.

PN W09603140-A1.

PD 08-FEB-1996.

PR 21-JUL-1995; U09234.

PR 21-JUL-1994; US-278634.

PR 23-NOV-1994; US-344824.

PR 30-MAY-1995; US-452843.

PA (CYTE-) CYTEL CORP.

PI Sette A, Sidney J.

DR WPI; 96-116784/12.

PT Composn. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications

PS Claim 2: Page 26: 32pp; English.

CC The sequences given in R89362-82 are immunogenic peptides which were used in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which allows binding of more than one HLA molecule. It pref. comprises two conserved residues, a first at the 2nd position from the N-terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also useful in compositions for in vivo and ex vivo therapeutic and diagnostic applications, e.g. the treatment of cancer and viral infections, e.g. hepatitis B and C.

CC Sequence 9 AA.

SO

Query Match 74.0%; Score 57; DB 18; Length 9;

Best Local Similarity 77.8%; Pred. No. 2.02e+01;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1 fpylrvsl 9

OY 1 fpylrvsl 9

RESULT 3

ID W38927 standard; peptide; 15 AA.

AC W38927;

DT 27-MAR-1998 (first entry)

DE Peptide resembling an SH3 domain binding peptide SEQ ID NO:324.

KW Cortactin; SH3 domain; binding peptide; Src homology region 3;

KW tyrosine kinase; immune response; lymphokine; interleukin 1; Nck;

KW Abl; PLCgamma; p53bp2; Crk; Yes; Grb2.

OS Synthetic.

PN W09730074-A1.

PD 21-AUG-1997.

PR 14-FEB-1997; U02298.

PR 16-FEB-1996; US-602999.

PA (CYTO-) CYTOGEN CORP.

PA (UYNC-) UNIV NORTH CAROLINA.

PI Der CJ, Fowlkes DM, Kay BK, Quilliam LA, Rider JE.

DR WPI; 97-424972/39.

PT Src homology region 3 binding peptide - used to activate Src tyrosine kinase(s) and to stimulate immune response by increasing production of certain lymphokine(s), e.g. interleukin-1

PS Claim 22: Page 90; 131pp; English.

CC The present sequence represents a peptide which resembles a Src homology region 3 (SH3) binding peptide. SH3 binding peptides are selected from: (a) peptides which bind the SH3 domain of Cortactin; (b) peptides which bind the middle SH3 domain of Nck; (c) peptides which bind the SH3 domain of Abl; (d) peptides which bind the SH3 domain of Src; (e) peptides which bind the SH3 domain of PLC gamma; (f) peptides which bind the SH3 domain of p53bp2; (g) peptides which bind the amino-terminal SH3

CC domain of Crk; (h) peptides which bind the SH3 domain of Yes; and (i) peptides which bind the amino-terminal SH3 domain of Grb2. The purified CC binding peptides can be used in the method to identify inhibitors of CC their binding to their respective SH3 domains, which could be used to CC modulate the pharmacological activity of proteins or polypeptide CC containing the SH3 domain. The peptides can also be used to activate CC Src or Src-related protein tyrosine kinases, to stimulate the immune CC response by increasing the production of certain lymphokines, e.g. CC tumour necrosis factor-alpha and interleukin-1, or to deliver a CC conjugated molecule to certain cellular compartments containing Src or CC Src related proteins.

CC Sequence 15 AA;

SO

Query Match 72.7%; Score 56; DB 26; Length 15;

Best Local Similarity 85.7%; Pred. No. 2.59e+01;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 fpylrvsl 7

OY 3 fpylrvsl 9

RESULT 4

ID W00183 standard; Protein; 516 AA.

AC W00183;

DT 18-OCT-1996 (first entry)

DE Cytochrome P4501A2.

KW Primer; polymerase chain reaction; PCR; amplify cytochrome P4501A2;

KW human; antibody; detection.

OS Synthetic.

PN J08143600-A.

PD 04-JUN-1996.

PR 14-NOV-1994; 279537.

PR 14-NOV-1994; JP-279537.

PA (SUMO) SUMITOMO CHEM CO LTD.

DR WPI; 96-318961/32.

PT Antibody recognising human-originated cytochrome P4501A2 - specifically recognises the species of P450 cytochrome present in a sample

PS Example 1: Page 11-13; 15pp; Japanese.

CC This sequence represents the cytochrome P4501A2. The cDNA sequence was amplified using the primers given in T33311-12. The human derived CC cytochrome P4501A2 was used in the generation of an antibody which CC is specific for this type of cytochrome. These antibodies may be CC used in the rapid and accurate determination of the exact cytochrome CC species present in a sample.

CC Sequence 516 AA;

SO

Query Match 72.7%; Score 56; DB 18; Length 516;

Best Local Similarity 85.7%; Pred. No. 2.59e+01;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 239 fpylrvsl 245

OY 3 fpylrvsl 9

RESULT 5

ID R93167 standard; Protein; 516 AA.

AC R93167;

DT 11-OCT-1996 (first entry)

DE Human cytochrome P450 molecular species 1A2 protein.

KW Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;

KW liver; yeast; expression vector; NADPH-P450 reductase; Adh gene promoter;

KW evaluation; safety; fusion protein; metabolite; detoxification;

KW carcinogenic.

OS Homo sapiens.

PN J08056695-A.

PD 05-MAR-1996.

PR 15-JUL-1994; 164184.

PR 20-JUL-1993; JP-201120.

PR 30-JUL-1993; JP-208279.

PR 17-JUN-1994; JP-136053.  
 PA (SUMO) SUMITOMO CHEM CO LTD.  
 DR MPI: 96-182311/19.  
 DR N-PSDB: T28380.  
 PT Novel method for the evaluation of the safety of a cpd. - using a  
 PT human cytochrome P450 and yeast NADPH reductase to determine whether  
 PT the analyte cpd. is detoxified or metabolised to a carcinogen  
 PS Example 1: Page 18-20; 74pp; Japanese.  
 CC This is the amino acid sequence of the human cytochrome P450 molecular  
 CC species 1A2 protein. The corresp. 1.5 kb fragment encoding the protein  
 CC was amplified from a human liver derived cDNA library using primers  
 CC T26923-4. The prod. was cloned into the yeast expression vector pAH5N  
 CC to generate plasmid p1A2 for prodn. of the cytochrome only or into the  
 CC vector pAHRR to generate the plasmid p1A2R for co-prodn. with the yeast  
 CC NADPH-P450 reductase. The sequence is placed under control of the yeast  
 CC ADH gene promoter and terminator.  
 CC The vectors are used in a method for evaluating the safety of a cpd. by  
 CC reacting the test cpd. with recombinantly produced human cytochrome P450  
 CC mol. species 1A2, 2C9 (T28381), 2E1 (T28382), 3A4 (T28383) or their  
 CC variants (T28384-98) together with yeast NADPH-P450 reductase (either as  
 CC a fused protein or as a cell extract) and analysing the resultant  
 CC metabolite. The cpd. is considered "safer" if it is detoxified or not  
 CC rendered carcinogenic or "unsafe" if it is not detoxified or is  
 CC metabolised to a carcinogenic cpd.  
 SQ Sequence 516 AA;

Query Match 72.7%; Score 56; DB 18; Length 516;  
 Best Local Similarity 85.7%; Pred. No. 2.59e+01;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 239 fplrryl 245  
 |||:||||  
 QY 3 FPIVRYL 9

RESULT 6  
 ID R72360 standard; Protein: 516 AA.  
 AC R72360;  
 DT 10-NOV-1995 (first entry)  
 DE Human cytochrome P450 molecular species 1A2 protein.  
 KW Human cytochrome P450; amplification; PCR; primer; expression vector;  
 KM yeast NADPH-P450 reductase; safety; fusion protein; metabolite;  
 OS Carcinogen; mutagen; liver metabolism.  
 OS Homo sapiens.  
 PN EP-644267-A.  
 PD 22-MAR-1995.  
 PF 20-JUL-1994; 111298.  
 PR 20-JUL-1993; JP-201120.  
 PR 21-JUL-1993; JP-180246.  
 PR 30-JUL-1993; JP-208279.  
 PA (HAYA/) HAYASHI K  
 PA (SUMO) SUMITOMO CHEM CO LTD.  
 PI Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;  
 PI Yabusaki Y;  
 DR MPI: 95-116991/16.  
 DR N-PSDB: 087714.  
 PT Evaluation of safety of a chemical cpd. - using recombinant yeast  
 PT expressing human cytochrome P450 and a yeast NADPH-P450 reductase  
 PS Examples; Page 18-21; 124pp; English.  
 CC The amino acid sequence of the human cytochrome P450 species 1A2. The  
 CC 1.5 kb cDNA was amplified by PCR using the primers 087733-4. The product  
 CC was cloned into the yeast expression vectors pAH5N or pAHRR to produce  
 CC the vectors p1A2 for the expression of the cytochrome P450 alone or p1A2R  
 CC co-expressed with the yeast NADPH-P450 reductase, respectively.  
 CC The vectors are used in a method for evaluating the safety of a chemical  
 CC compound by reacting the chemical compound with recombinantly produced  
 CC human cytochrome P450 molecular species 1A2, 2C9 (087715), 2E1 (087716)  
 CC or 3A4 (087717) or their auxiliary species and variants (087718-32) and  
 CC yeast NADPH-P450 reductase, either as a fused protein or in cell  
 CC extracts, and analysing the resulting metabolite to assess the safety of  
 CC the chemical compound. The method is useful for determining whether the  
 CC chemical compound, or its metabolite, will be converted into a  
 CC carcinogenic or mutagenic form through metabolism in the liver.

SQ Sequence 516 AA;  
 Query Match 72.7%; Score 56; DB 13; Length 516;  
 Best Local Similarity 85.7%; Pred. No. 2.59e+01;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 239 fplrryl 245  
 |||:||||  
 QY 3 FPIVRYL 9

RESULT 7  
 ID R60101 standard; Protein: 713 AA.  
 AC R60101;  
 DT 15-MAR-1995 (first entry)  
 DE Canine zona pellucida Czp2.  
 KW Canine; dog; zona pellucida; ZP; Czp2; contraceptive; vaccine;  
 KM antigen.  
 OS Canis familiaris.  
 PN J06189766-A.  
 PD 12-JUL-1994.  
 PF 25-DEC-1992; 359265.  
 PR 25-DEC-1992; JP-359265.  
 PA (TOFU) TONEN CORP.  
 DR MPI: 94-259553/32.  
 DR N-PSDB: Q70072.  
 PT New DNA sequence encoding canine zona pellucida Czp2 - useful for  
 PT the prodn. of a canine contraceptive vaccine antigen  
 PS Claim 1: Page 8-10; 10pp; Japanese.  
 CC The Czp2 DNA (Q70072) was prepd. by the cloning of Czp2(75-520) -  
 CC Q81700 using the primers given in Q70073-74, Czp2(1-65) - Q81804  
 CC using the primers given in Q70082-83, Czp2(42-103) - Q81803 using  
 CC the primers given in Q70079-81 and Czp2(487-713) - Q81957 using the  
 CC primers given in Q70075-78.  
 SQ Sequence 713 AA;

Query Match 72.7%; Score 56; DB 11; Length 713;  
 Best Local Similarity 71.4%; Pred. No. 2.59e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 509 ypvrryl 515  
 :||:||||  
 QY 3 FPIVRYL 9

RESULT 8  
 ID R5198 standard; Protein: 715 AA.  
 AC R5198;  
 DT 31-JAN-1995 (first entry)  
 DE Canine zona pellucida ZPA protein.  
 KW Dog; canine; zona pellucida; ZPA; immunoreception.  
 OS Canis familiaris.  
 OS Key location/Qualifiers  
 FT protein 1..715  
 FT /label= canine\_ZPA  
 PN M09411019-A.  
 PD 26-MAY-1994.  
 PF 06-NOV-1993; U10851.  
 PR 09-NOV-1992; US-973341.  
 PR 29-JAN-1993; US-012990.  
 PA (ZONA-) ZONAGEN INC.  
 PI Harris JD, Hsu KT, Podolski JS;  
 DR MPI: 94-183156/22.  
 DR N-PSDB: Q65608.  
 PT Use of zona pellucida proteins and antibodies - for inducing  
 PT reproducible transient infertility or permanent sterility in  
 PT female mammals  
 PS Claim 40; Page 88-90; 154pp; English.  
 CC A commercially available 16 week old canine ovarian cDNA expression  
 CC library in lambda gtl was screened using antibodies raised against  
 CC heat solubilised canine zona pellucida. The largest candidate clone  
 CC was used to rescreen the library and to isolate clones which were  
 CC used as probes in Southern hybridisations. Sequences coding for

CC canine ZPA and ZPC proteins were obtained (Q65608 and Q65609,  
CC respectively). R55198 is the deduced amino acid sequence for ZPA.  
SQ Sequence 715 AA;

Query Match 72.7% Score 56; DB 10; Length 715;  
Best Local Similarity 71.4% Pred. No. 2.59e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 511 ypvrvy1 517

1:|||||  
QY 3 FPIVRYL 9

# RESULT 9

ID R60532 standard; Protein; 716 AA.  
AC R60532;  
DT 12-MAY-1995 (first entry)  
DE Feline zona pellucida (FZP)-2.  
KW Feline zona pellucida; FZP-2; contraceptive vaccine antigen.  
OS Felis domesticus.  
PN J06217777-A.  
PD 09-AUG-1994.  
PF 29-JAN-1993; 013496.  
PR 29-JAN-1993; JP-013496.  
PA (TOFU) TONEN CORP.  
DR WPI; 94-305384/38.  
DR N-PSDB; 071287.  
PT New DNA sequence coding feline zona pellucida (Fzp) 2 - used for  
PS prodn. of contraceptive vaccine antigen for cats  
The feline zona pellucida (Fzp)-2 coding sequence was obtained by  
PCR amplification of cDNA synthesised from mRNA isolated from cat  
ovaries. Subclonings of the 716 amino acid Fzp-2 protein can be  
expressed for use as antigens in contraceptive vaccines for cats.  
SQ Sequence 716 AA;

Query Match 72.7% Score 56; DB 12; Length 716;  
Best Local Similarity 71.4% Pred. No. 2.59e+01;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 513 ypvrvy1 519

1:|||||  
QY 3 FPIVRYL 9

# RESULT 10

ID R55200 standard; Protein; 716 AA.  
AC R55200;  
DT 01-FEB-1995 (first entry)  
DE Feline zona pellucida ZPA protein.  
KW Cat; feline; zona pellucida; ZPA; immunoreception.  
OS Felis domesticus.  
FH Key: \* Location/Qualifiers  
FT protein 1..716  
FT /label= feline\_ZPA  
PN W09411019-A.  
PD 26-MAY-1994.  
PE 06-NOV-1993; U10851.  
PR 09-NOV-1992; US-973341.  
PR 29-JAN-1993; US-012990.  
PA (ZONA-) ZONAGEN INC.  
PA Harris JD, Hsu KT, Podolski JS;  
PI WPI; 94-183156/22.  
DR N-PSDB; 065610.  
PT Use of zona pellucida proteins and antibodies - for inducing  
PT reproducible transient infertility or permanent sterility in  
PT female mammals  
PS Claim 40; Page 97-99; 154pp; English.  
CC A cDNA library was prepared in lambda gt10 from mRNA isolated from  
CC ovaries of 3-4 month old cats. Plaques were screened using a  
CC mixture of probes encoding porcine ZPA, ZPB and ZPC proteins.  
CC Positive clones were analysed further by Southern hybridisation  
CC using the porcine probes and clones encoding feline ZPA, ZPB and

CC ZPC proteins were identified. The deduced amino acid sequence  
CC (R55200) from the feline ZPA clone was approximately 75% homologous  
CC to canine ZPA protein.  
SQ Sequence 716 AA;

Query Match 72.7% Score 56; DB 10; Length 716;  
Best Local Similarity 71.4% Pred. No. 2.59e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 513 ypvrvy1 519

1:|||||  
QY 3 FPIVRYL 9

# RESULT 11

ID R89369 standard; peptide; 9 AA.  
AC R89369;  
DT 18-SEP-1996 (first entry)  
DE Cw6 consensus peptide derived immunogenic peptide #1.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN W09603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PA Sette A, Sidney J;  
DR WPI; 96-116784/12.  
PT Compr. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
DE The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g. the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 68.8% Score 53; DB 18; Length 9;  
Best Local Similarity 85.7% Pred. No. 5.40e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 lpyprvr 7

1:|||||  
QY 1 IPIPIVR 7

# RESULT 12

ID R06998 standard; protein; 713 AA.  
AC R06998;  
DT 18-JAN-1991 (first entry)  
DE Mouse ZP2 protein exhibiting epitope for Ab which inhibits fertilisation  
DE of an oocyte by a sperm.  
KW Contraceptive; ZP3 protein; zona pellucida.  
OS Mus musculus.  
PN US7364379-A.  
PD 28-AUG-1990.  
PF 12-JUN-1989; 364379.  
PR 12-JUN-1989; US-364379.  
PA (USSH) NAT INST DIABETES.  
PA Jurrien D;  
DR WPI; 90-297734/39.  
DR NP-PSDB; 006005.

PT Contraceptive antibody vaccine for mammalian female - comprises  
 PT peptide epitope of zona pellucida protein, minimises possibility  
 PT of birth defects if failed contraception.  
 PS Disclosure: Fig 3: 93pp: English.  
 CC Vaccine provides long term, non-permanent contraception in mammals,  
 CC by inhibition of fertilisation rather than abortive methods, thus  
 CC minimising risk of birth defects.  
 CC Gene product comprises epitope to zona pellucida protein and vectors  
 CC and transformed expression systems are also claimed.  
 SQ Sequence 713 AA;

Query Match 68.8%; Score 53; DB 2; Length 713;  
 Best Local Similarity 71.4%; Pred. No. 5.40e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 507 YPIVRYL 513  
 :||:||||  
 QY 3 FPIVRYL 9

RESULT 13  
 ID W19612 standard; Protein: 271 AA.  
 AC W19612;  
 DT 22-AUG-1997 (first entry)  
 DE Human growth hormone secretagogue receptor type I.  
 KW Growth hormone secretagogue receptor; GHSR;  
 KW G protein coupled receptor.  
 OS Homo sapiens.  
 PN WO9722004-A1.  
 PD 19-JUN-1997.  
 PF 10-DEC-1996; U194442.  
 PR 13-DEC-1995; US-008584.  
 PR 06-JUN-1996; US-019259.  
 PA (MERL) MERCK & CO INC.  
 PI Chang L, Feiginger SD, Howard AD, Pong SS, Van Der Ploeg L;  
 DR WPI: 97-332924/30.  
 DR N-PSDB: T68666.  
 PT Determining presence of G protein cell membrane receptor and ligands  
 PT - requires presence of G protein subunit  
 PS Example 10; Fig 12: 77pp: English.  
 CC Human type I growth hormone secretagogue receptor (GHSR) (W19612)  
 CC is a novel member of the 7-transmembrane domain-containing  
 CC G-protein linked receptor superfamily. Its amino acid sequence was  
 CC deduced from a cDNA clone (T68666) isolated from a human pituitary  
 CC library using swine type I GHSR clone 7-3 (T68662) as probe.  
 CC Clone 7-3 had been identified using a novel assay for the detection  
 CC of nucleic acids encoding G protein coupled receptors. Other  
 CC human, pig and rat GHSRs (see also W19808-11 and W19613) have  
 CC also been identified. GHSRs can be used in assays to determine the  
 CC presence of a growth hormone secretagogue (GHS) or to determine the  
 CC concentration of GHS in body fluids. They can also be used to  
 CC amplify the effect of GHS administered to a patient by providing  
 CC increased downstream signal.  
 SQ Sequence 271 AA;

Query Match 66.2%; Score 51; DB 23; Length 271;  
 Best Local Similarity 55.6%; Pred. No. 8.76e+01;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 182 lpfhvryl 190  
 :||:||||  
 QY 1 IPEPIVRYL 9

RESULT 14  
 ID W19219 standard; Protein: 271 AA.  
 AC W19219;  
 DT 09-SEP-1997 (first entry)  
 DE Human growth hormone secretagogue receptor type I.  
 KW Growth hormone secretagogue receptor; GHSR;  
 KW G protein coupled receptor; ss.  
 OS Homo sapiens.  
 PN WO9721730-A1

PD 19-JUN-1997.  
 PF 10-DEC-1996; U19445.  
 PR 13-DEC-1995; US-008582.  
 PR 06-JUN-1996; US-018962.  
 PA (MERL) MERCK & CO INC.  
 PI Arena JP, Cully DF, Feiginger SD, Howard AD, Liberator PA;  
 PI Scheeffler JM, Van Der Ploeg L;  
 DR WPI: 97-332725/30.  
 DR N-PSDB: T69758.  
 PT Receptor of growth hormone receptor family - specifically  
 PT secretagogue or secretagogue like receptor, useful to screen for  
 PT specific binding agents for growth hormone deficiency treatment  
 PS Claim 12; Fig 12: 70pp: English.  
 CC Human type I growth hormone secretagogue receptor (GHSR) (W19219)  
 CC is a novel member of the 7-transmembrane domain-containing  
 CC G-protein linked receptor superfamily. Its amino acid sequence was  
 CC deduced from a cDNA clone (T69758) isolated from a human pituitary  
 CC library. Other human, pig and rat GHSRs (see also W19215-18,  
 CC W19220) have also been identified. GHSRs can be used to screen  
 CC for specific binding agents useful in treatment of conditions  
 CC related to growth hormone shortage, e.g. in growth hormone  
 CC deficient children, and elderly patients with musculoskeletal  
 CC impairment and recovering from hip fracture and osteoporosis.  
 SQ Sequence 271 AA;

Query Match 66.2%; Score 51; DB 23; Length 271;  
 Best Local Similarity 55.6%; Pred. No. 8.76e+01;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 182 lpfhvryl 190  
 :||:||||  
 QY 1 IPEPIVRYL 9

RESULT 15  
 ID R87029 standard; Protein: 341 AA.  
 AC R87029;  
 DT 31-JUL-1996 (first entry)  
 DE Varicella zoster virus thymidine kinase protein.  
 KW Varicella zoster virus; thymidine kinase; open reading frame; TARA box;  
 KW polyA signal; chimaeric construct; enzyme; toxin; target; albumin;  
 KW tissue specific regulatory sequence; transcription; produg; cell death;  
 KW hepatocellular cancer; arabinoside; carboxypeptidase; cytosine deaminase;  
 KW alkaline phosphatase; penicillin V-amide.  
 OS Varicella-zoster virus.  
 PN EP-690129-A1.  
 PD 03-JAN-1996.  
 PF 29-AUG-1990; 111593.  
 PR 30-AUG-1989; GB-019607.  
 PA (WELL) WELLCOME FOUND LTD.  
 PI Huber B, Krenitsky TA, Richards CA;  
 DR WPI: 96-051263/06.  
 DR N-PSDB: T07314.  
 PT Use of DNA chimera and produg for cancer therapy - where chimera  
 PT expresses produg-converting enzyme in cancer cells to form  
 PT cytotoxic agent  
 PS Example 1; Fig 2b: 30pp: English.  
 CC This is the amino acid sequence of the Varicella zoster virus (VZV)  
 CC thymidine kinase (TK). The gene corresp. to nucleotides 64276-66255 of  
 CC the complete VZV genome. The sequence contains the open reading frame and  
 CC the 5' sequence (including the TARA box) and the 3' sequence (including  
 CC a polyA signal) of the VZV gene. The AccI-KerI fragment (bases 519-1900)  
 CC can be used to generate chimaeric constructs comprising a gene  
 CC encoding an enzyme able to generate a cpd. toxic to a cancer cell,  
 CC under control of a target tissue specific regulatory sequence e.g.  
 CC ALB, the albumin transcriptional regulatory sequence. The chimeric  
 CC sequences can be administered by transfection, electroporation,  
 CC microinjection, liposomal transfer, ballistic barrage or by retroviral  
 CC infection and are partic. useful for treating hepatocellular cancer.  
 CC The enzyme is administered with the produg. In this case, e.g. a  
 CC purine or pyrimidine arabinoside esp. 9-beta-arabinoarabosyl-6-methoxy-  
 CC 9H-purine. The produg is then converted to a toxic cpd. leading to cell  
 CC death. Other examples of enzymes and produgs include carboxypeptidase G2

CC and p-N,N-bis(2-chloroethyl)aminobenzoyl-glutamic acid; alkaline  
phosphatase and etoposide phosphate, doxorubicin phosphate or mitomycin  
phosphate; penicillin-V-amides and a phenoxacetamide deriv. of  
CC doxorubicin or melphalan or cytosine deaminase and 5-fluorocytosine.  
SQ Sequence 341 AA;

Query Match 66.2%; Score 51; DB 17; Length 341;  
Best Local Similarity 66.7%; Pred. No. 8.76e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
DB 137 [cfdp] 145  
QY 1-PPPIVRL 9

Search completed: Fri Sep 11 13:13:39 1998  
Job time : 15 secs.

(TM)

**MPSrch\_pp** protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Fri Sep 11 13:14:21 1998; MasPar time 3.35 Seconds
```

Tabular output not generated.

Title: >US-08-452-843-10  
Description: (1-9) from US08452843.pep  
Perfect Score: 77  
Sequence: 1 IPEPIVRL 9

Scoring table: PAM 150

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: pir56

Statistics: Mean 24.683; Variance 37.209; scale 0.6663

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| No. | Score | Query Match | Length | DB | ID     | Description           | Pred. No. |
|-----|-------|-------------|--------|----|--------|-----------------------|-----------|
| 1   | 57    | 74.0        | 492    | 2  | A69715 | spore formation prote | 3.67e+00  |
| 2   | 56    | 72.7        | 513    | 2  | UX0190 | cytochrome P450 1A2 - | 5.49e+00  |
| 3   | 56    | 72.7        | 515    | 1  | O4RHU4 | cytochrome P450 1A2 - | 5.49e+00  |
| 4   | 56    | 72.7        | 516    | 1  | O4RBN  | cytochrome P450 1A2 - | 5.49e+00  |
| 5   | 56    | 72.7        | 524    | 2  | C37222 | cytochrome P450 1A1,  | 5.49e+00  |
| 6   | 56    | 72.7        | 715    | 2  | S70397 | zona pellucida glycop | 5.49e+00  |
| 7   | 56    | 72.7        | 716    | 2  | S70398 | zona pellucida glycop | 5.49e+00  |
| 8   | 55    | 71.4        | 212    | 2  | F69440 | hypothetical protein  | 8.17e+00  |
| 9   | 55    | 71.4        | 310    | 1  | K19ETH | thymidine kinase (EC  | 8.17e+00  |
| 10  | 55    | 71.4        | 323    | 2  | A70029 | hypothetical protein  | 8.17e+00  |
| 11  | 55    | 71.4        | 350    | 1  | K19EFC | thymidine kinase (EC  | 8.17e+00  |
| 12  | 55    | 71.4        | 452    | 2  | I52373 | gastric inhibitory po | 8.17e+00  |
| 13  | 55    | 71.4        | 465    | 2  | JC2462 | gastric inhibitory po | 8.17e+00  |
| 14  | 55    | 71.4        | 466    | 2  | S66676 | glucose-dependent ins | 8.17e+00  |
| 15  | 55    | 71.4        | 466    | 2  | G05234 | gastric inhibitory po | 8.17e+00  |
| 16  | 55    | 71.4        | 491    | 2  | I37411 | glucose-dependent ins | 8.17e+00  |
| 17  | 55    | 71.4        | 1055   | 2  | I59531 | thyrotropin-releasing | 8.17e+00  |
| 18  | 54    | 70.1        | 543    | 2  | A61400 | cytochrome P450 1A2 - | 1.21e+01  |
| 19  | 54    | 70.1        | 513    | 1  | O4MSM3 | acetanilide 4-hydroxy | 1.21e+01  |
| 20  | 53    | 68.8        | 189    | 2  | S43558 | membrane protein B028 | 1.79e+01  |
| 21  | 53    | 68.8        | 676    | 2  | A45984 | sperm-binding glycopr | 1.79e+01  |
| 22  | 53    | 68.8        | 713    | 2  | A34782 | sperm-binding glycopr | 1.79e+01  |
| 23  | 52    | 67.5        | 236    | 2  | H65118 | hypothetical adenine- | 2.62e+01  |

|    |    |      |       |   |        |                       |           |
|----|----|------|-------|---|--------|-----------------------|-----------|
| 45 | 50 | 64.9 | 84.5  | 1 | RKXSDA | RNA-directed RNA poly | 5.56e+01  |
| 44 | 50 | 64.9 | 84.4  | 1 | RKXSPD | SECL protein - yeast  | 5.56e+01  |
| 43 | 50 | 64.9 | 72.4  | 2 | S17479 | RNA-directed RNA poly | 5.56e+01  |
| 42 | 50 | 64.9 | 48.8  | 2 | D64108 | glucanase transport r | 5.56e+01  |
| 41 | 50 | 64.9 | 38.2  | 2 | G01589 | ionizing radiation re | 5.56e+01  |
| 40 | 50 | 64.9 | 26.1  | 2 | S45050 | biopolymer transport  | 5.56e+01  |
| 39 | 50 | 64.9 | 22.7  | 2 | S74918 | Diopolymer transport  | 5.56e+01  |
| 38 | 50 | 64.9 | 13.9  | 2 | S78036 | endocuticular protein | 5.56e+01  |
| 37 | 51 | 66.2 | 71.2  | 2 | S70434 | zona pellucida glycop | 3.83e+01  |
| 36 | 51 | 66.2 | 61.2  | 2 | A34967 | sterol esterase (EC 3 | 3.83e+01) |
| 35 | 51 | 66.2 | 56.3  | 2 | A70038 | L-lactate permease ho | 3.83e+01  |
| 34 | 51 | 66.2 | 46.3  | 2 | A46172 | glucagon-like peptide | 3.83e+01  |
| 33 | 51 | 66.2 | 38.0  | 2 | VQ2338 | omega-3 fatty acid de | 3.83e+01  |
| 32 | 51 | 66.2 | 34.1  | 1 | K1BE36 | thymidine kinase (EC  | 3.83e+01) |
| 31 | 51 | 66.2 | 34.1  | 1 | K1BE37 | thymidine kinase (EC  | 3.83e+01) |
| 30 | 51 | 66.2 | 34.1  | 1 | K1BE40 | thymidine kinase (EC  | 3.83e+01) |
| 29 | 51 | 66.2 | 34.1  | 1 | K1BE40 | thymidine kinase (EC  | 3.83e+01) |
| 28 | 51 | 66.2 | 27.5  | 2 | S40005 | trypsin (EC 3.4.21.4) | 3.83e+01  |
| 27 | 51 | 66.2 | 26.1  | 2 | F64305 | hypothetical protein  | 3.83e+01  |
| 26 | 51 | 66.2 | 26.1  | 2 | F64305 | hypothetical protein  | 3.83e+01  |
| 25 | 51 | 66.2 | 24.2  | 2 | G64498 | phosphoribosylaminoi  | 3.83e+01  |
| 24 | 52 | 67.5 | 106.4 | 2 | SS2667 | serine/threonine-spec | 2.62e+01  |

## ALIGNMENTS

| RESULT ENTRY   | 1   |                          |
|--|---|--------------------------|
| TITLE  |   |                          |
| ALTERNATE_NAMES  |   |                          |
| ORGANISM   |   |                          |
| DATE   |   |                          |
| ACCESSIONS   |   |                          |
| REFERENCE  |   |                          |
| #authors   |   |                          |
| A69715   | #type complete  |                          |
|  | spore formation protein SPOVAF - <i>Bacillus subtilis</i> |                          |
|  | stage V sporulation protein AF                            |                          |
|  | #format_name <i>Bacillus subtilis</i>                     |                          |
|  | 05-Dec-1997 #sequence_revision                            | 05-Dec-1997 #text_change |
|  | 13-Feb-1998   |                          |
| A69715; S45533; PS0430   |   |                          |
| A69580   |   |                          |
| Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bartolo, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capanu, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, J.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devigne, K.M.; Duesternhoef, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Fougier, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallitz, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Goldightly, E.J.; Grandi, G.; Giuseppe, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koelter, P.; Konungstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Labber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maasda, S.; Mauviel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivoletta, C.; Roche, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Seror, P.; Shin, B.S.; Solido, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terststra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandendool, M.; Vannier, F.; Vassartoli, A.; Vairi, A.; Wamplitz, R.; Wedler, E.; Wedler, H.; Weitzneger, T.; Winers, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A. |   |                          |
| Nature (1997) 390:249-256  |   |                          |
| The complete genome sequence of the Gram-positive bacterium <i>Bacillus subtilis</i> .   |   |                          |

```

#accession      A69715
#status         nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues       1-492 ##label KUN
#experimental_source strain 168
REFERENCE
#authors        S45533
               Sotokin, A.; Zumstein, E.; Azevedo, V.; Ehrlich, S.D.;
               Serior, P.
#submission     submitted to the EMBL Data Library, November 1993
#accession      S45533
#molecule_type DNA
#residues       81-431, 'L', 433-492 ##label SCR
#cross-references EMBL:L09228; NID:g410114; PID:g410115
               JU0471
REFERENCE
#authors        Yamamoto, J.; Shimizu, M.; Yamane, K.
               Agric. Biol. Chem. (1991) 55:1615-1616
#journal        Molecular cloning and analysis of nucleotide sequence of the
               Bacillus subtilis lysa gene region using B. subtilis phage
               vectors and a multi-copy plasmid, pub110.
#accession      PS0430
#molecule_type DNA
#residues       223-431, 'L', 433-466, 'P', 468-492 ##label YAM
#note           the authors translated the codon TCA for residue 392 as
               Thr, AGT for residue 419 as Ser, TAT for residues 425
               and 437 as Thr, CTT for residue 432 as Ile, TAC for
               residue 433 as Thr, and AGC for residue 478 as Thr

GENETICS
#gene           spovAF
KEYWORDS        sporulation; transmembrane protein
SUMMARY         #length 492 #molecular-weight 55606 #checksum 2481

Query Match
Best Local Similarity 85.7%; Score 57; DB 2; Length 492;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 232 PFPIRY 238
Qy 2 PFPIRY 8

```

```

#journal        J. Biochem. (1990) 107:826-833
#title          Purification and characterization of two forms of 2,3,4,7,
               8-pentachlorodibenzofuran-inducible cytochrome P-450 in
               hamster liver
#cross-references M01D:90361684
#accession      PX0036
#molecule_type protein
#residues       2-19 ##label KOG
#experimental_source liver
COMMENT         Cytochrome P-450 I family consists of two members, IA1 and IA2.
               Both of them are inducible by 3-methylcholanthrene and 2,3,7,
               8-tetrachlorodibenzo-p-dioxin, but have different substrate
               specificities.

GENETICS
#gene           CYP1A2
CLASSIFICATION  superfamily cytochrome P450
KEYWORDS        chromoprotein; electron transfer; endoplasmic reticulum;
               heme; iron; monooxygenase; oxidoreductase; transmembrane
               protein
FEATURE
456             #binding-site heme iron (Cys) (axial ligand) #status
               predicted
SUMMARY         #length 513 #molecular-weight 58082 #checksum 3323

Query Match
Best Local Similarity 72.7%; Score 56; DB 2; Length 513;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 238 PFPIRY 244
Qy 3 PFPIRY 9

```

```

ENTRY          2
TITLE          JX0190 #type complete
ALTERNATE_NAMES cytochrome P450 IA2 - golden hamster
ORGANISM        #formal.name Mesocricetus auratus #common.name golden hamster
DATE            31-Mar-1992 #sequence-revision 31-Mar-1992 #text-change
               23-Jan-1998
ACCESSIONS     JX0190; S13885; PX0036
REFERENCE      JX0189
#authors       Sagami, I.; Ohmachi, T.; Fujii, H.; Kikuchi, H.; Watanabe, M.
#journal       J. Biochem. (1991) 110:641-647
#title        Hamster cytochrome P-450 IA gene family, P-450 IA1 and P-450
               IA2 in lung and liver: cDNA cloning and sequence analysis.
#cross-references M01D:92138673
#accession     JX0190
#molecule_type mRNA
#residues      1-513 ##label SAG
#experimental_source lung and liver, microsome
REFERENCE      S13884
#authors       Lai, T.S.; Chiang, J.Y.L.
#journal       Arch. Biochem. Biophys. (1990) 283:429-439
#title        Cloning and characterization of two major
               3-methylcholanthrene inducible hamster liver cytochrome
               P450s.
#cross-references M01D:9112759
#accession     S13885
#molecule_type mRNA
#residues      1-48, 'F', 50-51, 'MC', 54-252, 'GG', 255-325, 'W', 327-355, 'L',
               357-464, 'Q', 486-513 ##label LAI
#cross-references EMBL:M6387; NID:g191354; PID:g191355
               PX0036
#authors       Koga, N.; Ariyoshi, N.; Nakashima, H.; Yoshimura, H.

```

```

ENTRY          3
TITLE          O4H04 #type complete
ALTERNATE_NAMES cytochrome P450 IA2 - human
ORGANISM        #formal.name Homo sapiens #common.name man
DATE            28-Dec-1987 #sequence-revision 31-Mar-1992 #text-change
               20-Mar-1998
ACCESSIONS     S16718; A25892; A23585; S07373; S22433; A60881
REFERENCE      S16718
#authors       Ikeya, K.; Jaiswal, A.K.; Owens, R.A.; Jones, J.E.; Nebert,
               D.W.; Kimura, S.
#journal       Mol. Endocrinol. (1989) 3:1399-1408
#title        Human CYP1A2: sequence, gene structure, comparison with the
               mouse and rat orthologous gene, and differences in liver
               IA2 mRNA expression.
#cross-references M01D:90114205
#accession     S16718
#molecule_type DNA
#residues      1-515 ##label IKE
#cross-references EMBL:M31664
REFERENCE      A25892
#authors       Quattrochi, L.C.; Pendurthi, U.R.; Okino, S.T.; Potenza, C.;
               Tukey, R.H.
#journal       Proc. Natl. Acad. Sci. U.S.A. (1986) 83:6731-6735
#title        Human cytochrome P-450 4 mRNA and gene: part of a multigene
               family that contains Alu sequences in its mRNA.
#cross-references M01D:86313652
#accession     A25892
#molecule_type DNA
#residues      1-78, 'S', 80-510, 'LP', 512-515 ##label QUA
#cross-references EMBL:L00388
REFERENCE      A90953
#authors       Quattrochi, L.C.; Okino, S.T.; Pendurthi, U.R.; Tukey, R.H.
#journal       DNA (1985) 4:395-400
#title        Cloning and isolation of human cytochrome P-450 cDNAs
               homologous to dioxin-inducible rabbit mRNAs encoding P-450
               4 and P-450 6.
#cross-references M01D:86081170
#accession     A23585
#molecule_type mRNA

```



```

#residues 295-310, 'L', 312-449, 'M', 450, 'V', 452-485 ##label QU2
##cross-references GB:M2078; NID:g181351; PID:g553246
REFERENCE
#authors Jaiswal, A.K.; Nebert, D.W.; Gonzalez, F.J.
#journal Nucleic Acids Res. (1986) 14:6773-6774
#title Human P(3)450: cDNA and complete amino acid sequence.
#cross-references M01D:86312938
#accession S07373
#status translation not shown
#molecule-type mRNA
#residues 1-515 ##label JAI
##cross-references EMBL:Z00036; NID:g30338; PID:g30339
REFERENCE
#authors Jaiswal, A.K.; Nebert, D.W.; McBride, O.W.; Gonzalez, F.J.
#journal J. Exp. Pathol. (1987) 3:1-17
#title Human P(3)450: cDNA and complete protein sequence, repetitive
Alu sequences in the 3' nontranslated region, and
localization of gene to chromosome 15.
#cross-references M01D:88061719
#accession S22433
#status preliminary
#molecule-type mRNA
#residues 1-515 ##label JA2
##cross-references EMBL:M5053; NID:g181307; PID:g181308
REFERENCE
#authors Wrighton, S.A.; Campanile, C.; Thomas, P.E.; Maines, S.L.;
Watkins, P.B.; Parker, G.; Mendez-Picon, G.; Hanlu, M.;
Shively, J.E.; Levin, W.; Guzelian, P.S.
#journal Mol. Pharmacol. (1986) 29:405-410
#title Identification of a human liver cytochrome P-450 homologous
to the major isoflavone-inducible cytochrome P-450 in the
rat
#accession A60881
#molecule-type protein
#residues 1-219 ##label MRI
GENETICS
#gene GDB:CY1A2
#cross-references GDB:118780; OMIM:124060
#map_position 15q22-15qter
#clonings 277/3; 318/1; 348/1; 389/2; 418/2
#superfamily cytochrome P450
#chromoprotein: electron transfer; endoplasmic reticulum;
heme; iron; microsome; monooxygenase; oxidoreductase;
transmembrane protein
FEATURE
458 #binding-site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY
#length 515 #molecular-weight 58294 #checksum 960
Query Match 72.7%; Score 56; DB 1; Length 515;
Best Local Similarity 85.7%; Pred. NO. 5.49e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 239 PIRATL 245
III:III
QY 3 PIRATL 9
RESULT 4
ENTRY #type complete
TITLE cytochrome P450 1A2 - rabbit
ALTERNATE_NAMES
acetanilide 4-hydroxylase (EC 1.14.14.-); cytochrome P450
LM4; cytochrome P450-4
ORGANISM
#formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 28-May-1986 #sequence-revision 24-Feb-1994 #text_change
30-Jan-1998
ACCESSIONS
B27821; S02038; B25143; A00187; A00188; A44250
A91910
#authors Kagawa, N.; Mihara, K.; Sato, R.
#journal J. Biochem. (1987) 101:1471-1479
#title Structural analysis of cloned cDNAs for polycyclic
hydrocarbon-inducible forms of rabbit liver microsomal

```

```

cytochrome P-450.
#cross-references M01D:88032911
#accession B27821
#molecule-type mRNA
#residues 1-120, 'H', 122-516 ##label KA2
##cross-references EMBL:X05686; NID:g1540; PID:g1541
REFERENCE
#authors Pompon, D.
#journal Eur. J. Biochem. (1988) 177:285-293
#title cDNA cloning and functional expression in yeast Saccharomyces
cerevisiae of beta-naphthoflavone-induced rabbit liver
P-450 LM4 and LM6.
#cross-references M01D:89052697
#accession S02038
#molecule-type mRNA
#residues 1-173, 'S', 175-207, 'H', 209-332, 'S', 234-298, 'G', 300-353,
'PG', 356-516 ##label POM
##cross-references EMBL:X13853; NID:g1532; PID:g1533
#note the authors translated the codon GAC for residue 276 as
Gln and CCC for residue 354 as Ala
REFERENCE
#authors Okino, S.T.; Quattrocchi, L.C.; Barnes, H.J.; Osanto, S.;
Griffin, K.J.; Johnson, E.F.; Tukey, R.H.
#journal Proc. Natl. Acad. Sci. U.S.A. (1985) 82:5310-5314
#title Cloning and characterization of cDNAs encoding 2,3,7,
8-tetrachlorodibenzo-p-dioxin-inducible rabbit mRNAs for
cytochrome P-450 isozymes 4 and 6.
#cross-references M01D:85270514
#accession B25143
#molecule-type mRNA
#residues 'H', 94-207, 'H', 209-287, 'I', 289-290, 'N', 292, 'MD', 295,
'MDGAHV', 303-308, 'T', 310-357, 'L', 359-461, 'I', 463-516
##label OKI
#cross-references EMBL:M11728; NID:g165578; PID:g165579
REFERENCE
#authors Ozols, J.
#journal J. Biol. Chem. (1986) 261:3965-3979
#title Complete amino acid sequence of a cytochrome P-450 isolated
from beta-naphthoflavone-induced rabbit liver microsomes.
Comparison with phenobarbital-induced and constitutive
isozymes and identification of invariant residues.
#cross-references M01D:86140205
#accession A00187
#molecule-type protein
#residues 2-21, 'S', 23-69, 'Q', 71-91, 'N', 93-171, 'F', 173-193, 'S',
195-207, 'FPOGM', 213-246, 'OPN', 250, 'R', 252-289, 'SH',
292-294, 296-298, 'G', 300-493, 'T', 495-516 ##label OZO
233-Ser and 247-Asn were also found
REFERENCE
#note A00188
#authors Fujita, V.S.; Black, S.D.; Tarr, G.E.; Koop, D.R.; Coon, M.J.
#journal Proc. Natl. Acad. Sci. U.S.A. (1984) 81:4260-4264
#title On the amino acid sequence of cytochrome P-450 isozyme 4 from
rabbit liver microsomes.
#cross-references M01D:84272618
#accession A00188
#molecule-type protein
#residues 2-45, 'S', 47, 'V', 49, 107-118, 133-173, 'S', 175-197, 'X',
199-206, 217-232, 'S', 234-241, 'V', 243-246, 255-264;
267-274, 297-298, 'G', 300-341, 362-376, 'XX', 379-381;
394-444, 'A', 446-477, 'X', 479-486, 500-511, 'S', 513, 'K'
##label FUJ
REFERENCE
#authors Yun, C.H.; Hammons, G.J.; Jones, G.; Martin, M.V.; Hopkins,
N.E.; Alworth, W.L.; Guengerich, F.P.
#journal Biochemistry (1992) 31:10556-10563
#title Modification of cytochrome P450 1A2 enzymes by the
mechanism-based inactivator 2-ethynylnaphthalene and the
photoaffinity label 4-azidobiphenyl.
#accession A44250
#molecule-type protein
#residues 176-185 ##label YUN
#note only this tryptic peptide was photoaffinity labeled by
4-azidobiphenyl, a substrate analog

```

COMMENT: There are three forms of this protein that differ only in the absence or presence of the first one or two residues.

GENETICS #gene  
CLASSIFICATION #superfamily cytochrome P450  
KEYWORDS #chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; monooxygenase; oxidoreductase; transmembrane protein

FEATURE 458  
#binding site heme iron (Cys) (axial ligand) #status predicted

SUMMARY #length 516 #molecular-weight 58334 #checksum 3338

Query Match 72.7%; Score 56; DB 1; Length 516;  
Best Local Similarity 85.7%; Pred. No. 5,49e+00;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 239 FPIRLYL 245  
111111  
QY 3 FPIVRYL 9

RESULT 5  
ENTRY C37222 #type fragment  
TITLE cytochrome P450 1A1, hepatic - dog (fragment)  
ALTERNATE\_NAMES cytochrome P450 (Dahl)  
ORGANISM #formal\_name Canis lupus familiaris #common\_name dog  
DATE 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 23-Jan-1998

ACCESSIONS C37222  
REFERENCE A37222  
#authors Uchida, T.; Komori, M.; Kitada, M.; Kamataki, T.  
#journal Mol. Pharmacol. (1990) 38:644-651  
#title Isolation of cDNAs coding for three different forms of liver microsomal cytochrome P-450 from polychlorinated biphenyl-treated beagle dogs.  
#cross-references MIMD:91042464

#accession C37222  
#status #not compared with conceptual translation  
#molecule\_type mRNA  
#residues 1-524 #label UCH

CLASSIFICATION #superfamily cytochrome P450  
KEYWORDS #chromoprotein; electron transfer; heme; iron; liver; monooxygenase; oxidoreductase; transmembrane protein

FEATURE 461  
#binding site heme iron (Cys) (axial ligand) #status predicted

SUMMARY #length 524 #checksum 6556  
Query Match 72.7%; Score 56; DB 2; Length 524;  
Best Local Similarity 85.7%; Pred. No. 5,49e+00;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 241 FPIRLYL 247  
111111  
QY 3 FPIVRYL 9

RESULT 6  
ENTRY S70397 #type complete  
TITLE zona pellucida glycoprotein A - dog  
ORGANISM #formal\_name Canis lupus familiaris #common\_name dog  
DATE 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 08-Sep-1997

ACCESSIONS S70397  
REFERENCE S70396  
#authors Harris, J.D.; Hibler, D.W.; Fontenot, G.K.; Hsu, K.T.; Yurewicz, E.C.; Sacco, A.G.

#journal DNA Seq. (1994) 4:361-393  
#title Cloning and characterization of zona pellucida genes and cDNAs from a variety of mammalian species: the ZPA, ZPB and ZPC gene families.  
#accession S70397

##status preliminary  
##molecule\_type mRNA  
##residues 1-715 #label HAR  
##cross-references EMBL:005779; NID:9458274; PID:9458275  
CLASSIFICATION #superfamily sperm-binding glycoprotein ZP2; ZP domain homology

FEATURE 368-628  
#domain ZP domain homology #label ZPH

SUMMARY #length 715 #molecular-weight 79938 #checksum 3009  
Query Match 72.7%; Score 56; DB 2; Length 715;  
Best Local Similarity 71.4%; Pred. No. 5,49e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 511 YPVVRYL 517  
111111  
QY 3 FPIVRYL 9

RESULT 7  
ENTRY S70398 #type complete  
TITLE zona pellucida glycoprotein A - cat  
ORGANISM #formal\_name Felis silvestris catus #common\_name domestic cat  
DATE 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 08-Sep-1997

ACCESSIONS S70398  
REFERENCE S70396  
#authors Harris, J.D.; Hibler, D.W.; Fontenot, G.K.; Hsu, K.T.; Yurewicz, E.C.; Sacco, A.G.  
#journal DNA Seq. (1994) 4:361-393  
#title Cloning and characterization of zona pellucida genes and cDNAs from a variety of mammalian species: the ZPA, ZPB and ZPC gene families.  
#cross-references EMBL:005776; NID:9458268; PID:9458269

#accession S70398  
#status preliminary  
#molecule\_type mRNA  
#residues 1-716 #label HAR  
##cross-references EMBL:005776; NID:9458268; PID:9458269  
CLASSIFICATION #superfamily sperm-binding glycoprotein ZP2; ZP domain homology

FEATURE 370-630  
#domain ZP domain homology #label ZPH

SUMMARY #length 716 #molecular-weight 80135 #checksum 6483  
Query Match 72.7%; Score 56; DB 2; Length 716;  
Best Local Similarity 71.4%; Pred. No. 5,49e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 513 YPVVRYL 519  
111111  
QY 3 FPIVRYL 9

RESULT 8  
ENTRY F64940 #type complete  
TITLE hypothetical protein bi798 - Escherichia coli (strain K-12)  
ORGANISM #formal\_name Escherichia coli  
DATE 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 14-Nov-1997

ACCESSIONS F64940  
REFERENCE A64720  
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

#journal Science (1997) 277:1453-1462  
#title The complete genome sequence of Escherichia coli K-12.  
#cross-references MIMD:97426617  
#accession F64940  
#status preliminary; nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA

```

##residues 1-212 ##label BLAT
##cross-references GB:AE000274; GB:U00096; NID:q1788089; PID:q1788099;
#experimental_source strain K-12, substrain MG1655
#length 212 #molecular-weight 23200 #checksum 2508
SUMMARY

Query Match 71.4%; Score 55; DB 2; Length 212;
Best Local Similarity 77.8%; Pred. No. 8.17e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 74 ILEFIVRYL 82
QY 1 IPEPIVRYL 9

RESULT 9
ENTRY KIBETH #type complete
TITLE thymidine kinase (EC 2.7.1.21) - turkey herpesvirus
ORANISM #formal_name turkey herpesvirus
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
05-Sep-1997
ACCESSIONS A33346
REFERENCE A33346
#authors Martin, S.L.; Aparicio, D.I.; Bandyopadhyay, P.K.
#journal J. Virol. (1989) 63:2847-2852
#title Genetic and biochemical characterization of the thymidine
#cross-references MUID:89259069
#accession A33346
#molecule_type DNA
#residues 1-310 #label MAR
#cross-references GB:M6659; NID:q330940; PID:q330941
CLASSIFICATION #superfamily herpesvirus thymidine kinase; herpesvirus
KEYWORDS ATP; DNA biosynthesis; P-loop; phosphotransferase
FEATURE 10-301
17-24 #domain herpesvirus thymidine kinase homology #label
117-121 #region nucleotide-binding motif A (P-loop) \
23 #binding_site ATP (lys) #status predicted
SUMMARY #length 310 #molecular-weight 35512 #checksum 7680

Query Match 71.4%; Score 55; DB 1; Length 310;
Best Local Similarity 85.7%; Pred. No. 8.17e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 131 FPIFRL 137
QY 3 FPIVRYL 9

RESULT 10
ENTRY A70029 #type complete
TITLE hypothetical protein yvax - Bacillus subtilis
ORANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Dec-1997
ACCESSIONS A70029
REFERENCE A69580
#authors Kinst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bartero, M.G.; Bessieres, P.;
Biolchini, A.; Borcher, S.; Boriss, R.; Boursier, L.; Brans,
A.; Brun, M.; Bricnell, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Enrich, S.D.; Emerson, P.T.; Entian, K.D.; Erilling, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Filtz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Goldlight, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Hage, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;

```

```

Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klier-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapius, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Nodack, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogihara, A.; Oudeg, B.; Park, S.H.; Parro,
V.; Pol, T.M.; Portetelle, D.; Porolik, S.; Prescott,
A.M.; Prescan, E.; Pulic, P.; Punelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serron, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takamaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Wetzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zimstein, E.;
Yoshikawa, H.; Zimstein, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#accession A70029
#status preliminary; nucleic acid sequence not shown;
translation not shown
#journal Nature
#title The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#accession A70029
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-323 #label KUN
#experimental_source strain 168
GENETICS
SUMMARY yvax #length 323 #molecular-weight 37450 #checksum 5488

Query Match 71.4%; Score 55; DB 2; Length 323;
Best Local Similarity 66.7%; Pred. No. 8.17e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 71 ILEFIVRYL 79
QY 1 IPEPIVRYL 9

RESULT 11
ENTRY KIBETH #type complete
TITLE thymidine kinase (EC 2.7.1.21) - turkey herpesvirus (strain
Fc-126)
ORANISM #formal_name turkey herpesvirus
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
28-Feb-1997
ACCESSIONS A33375
REFERENCE A33375
#authors Scott, S.D.; Ross, N.L.J.; Bins, M.M.
J. Gen. Virol. (1989) 70:3055-3065
Nucleotide and predicted amino acid sequences of the Marek's
disease virus and turkey herpesvirus thymidine kinase
genes; comparison with thymidine kinase genes of other
herpesviruses.
#cross-references MUID:90063552
#accession A33375
#molecule_type DNA
#residues 1-350 #label SCO
#cross-references EMBL:DO0561
CLASSIFICATION #superfamily herpesvirus thymidine kinase; herpesvirus
KEYWORDS ATP; DNA biosynthesis; P-loop; phosphotransferase
FEATURE 10-301
17-24 #domain herpesvirus thymidine kinase homology #label
117-121 #region nucleotide-binding motif A (P-loop) \
23 #binding_site ATP (lys) #status predicted

```

SUMMARY #length 350 #molecular-weight 39968 #checksum 1250

Query Match 71.4%: Score 55; DB 1; Length 350;

Best Local Similarity 85.7%: Pred. No. 8.17e+00; Mismatches 2; Indels 0; Gaps 0;

Db 131 PPIRYL 137  
111111  
3 PPIRYL 9

RESULT 12

ENTRY 153273 #type complete  
TITLE gastric inhibitory polypeptide receptor - rat  
ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
DATE 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 02-Aug-1996

ACCESSIONS 153273  
REFERENCE 153273

#authors Usdin, T.B.; Mezey, E.; Button, D.C.; Brownstein, M.J.; Bonner, T.I.

#journal Endocrinology (1993) 133:2861-2870  
#title Gastric inhibitory polypeptide receptor, a member of the secretin-vasoactive intestinal peptide receptor family, is widely distributed in peripheral organs and the brain.

#cross-references MUID:9406267

#accession 153273 preliminary; translated from GB/EMBL/DBJ

#status #molecule\_type mRNA

#residues 1-455 #label RES

SUMMARY #cross-references GB:L19660; NID:G431448; PID:G431449

Query Match 71.4%: Score 55; DB 2; Length 455;  
Best Local Similarity 77.8%: Pred. No. 8.17e+00; Mismatches 2; Indels 0; Gaps 0;

Db 269 IPWIVRYL 277  
111111  
1 IPWIVRYL 9

RESULT 13

ENTRY UC2462 #type complete  
TITLE gastric inhibitory polypeptide receptor - hamster  
ALTERNATE\_NAMES GIP receptor  
ORGANISM #formal\_name Cricetinae gen. sp. #common\_name hamster  
DATE 15-Feb-1995 #sequence\_revision 05-Apr-1995 #text\_change 10-Sep-1997

ACCESSIONS UC2462  
REFERENCE UC2462

#authors Yasuda, K.; Inagaki, N.; Yamada, Y.; Kubota, A.; Seino, S.; Seino, Y.

#journal Biochem. Biophys. Res. Commun. (1994) 205:1556-1562  
#title Hamster gastric inhibitory polypeptide receptor expressed in pancreatic islets and clonal insulin-secreting cells: its structure and functional properties.

#accession JC2462

FEATURE #molecule\_type mRNA  
#residues 1-462 #label YAS  
#cross-references DDBJ:D38103; NID:G644880; PID:G765087  
receptor; transmembrane protein

KEYWORDS 136-157 #domain transmembrane #status predicted #label TM1  
167-186 #domain transmembrane #status predicted #label TM2  
215-238 #domain transmembrane #status predicted #label TM3  
252-274 #domain transmembrane #status predicted #label TM4  
292-315 #domain transmembrane #status predicted #label TM5  
339-357 #domain transmembrane #status predicted #label TM6  
383-394 #domain transmembrane #status predicted #label TM7  
SUMMARY #length 462 #molecular-weight 52918 #checksum 9727

Query Match 71.4%: Score 55; DB 2; Length 462;

Best Local Similarity 77.8%: Pred. No. 8.17e+00; Mismatches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 269 IPWIVRYL 277  
111111  
1 IPWIVRYL 9

RESULT 14

ENTRY S66676 #type complete  
TITLE glucose-dependent insulinotropic protein receptor precursor - human

ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 13-Mar-1997

ACCESSIONS S66676  
REFERENCE S66676

#authors Voltz, A.; Goeke, R.; Lankat-Buttgereit, B.; Fehmann, H.C.; Bode, H.P.; Goeke, B.

#journal FEBS Lett. (1995) 373:23-29  
#title Molecular cloning, functional expression, and signal transduction of the GIP-receptor cloned from a human insulinoma.

#accession S66676

#status #molecule\_type mRNA

#residues 1-466 #label VOL

#cross-references GB:S79852

#note the authors translated the codon GCC for residue 427 as Leu

FEATURE 1-21

22-466 #domain signal sequence #status predicted #label SIG  
#product glucose-dependent insulinotropic protein receptor #status predicted #label MAT

SUMMARY #length 466 #molecular-weight 53142 #checksum 1170

Query Match 71.4%: Score 55; DB 2; Length 466;  
Best Local Similarity 77.8%: Pred. No. 8.17e+00; Mismatches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 272 IPWIVRYL 280  
111111  
1 IPWIVRYL 9

RESULT 15

ENTRY G02234 #type complete  
TITLE gastric inhibitory polypeptide receptor - human  
ALTERNATE\_NAMES GIP receptor  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 21-Dec-1990 #sequence\_revision 06-Jun-1997 #text\_change 15-Aug-1997

ACCESSIONS G02234  
REFERENCE G09336

#authors Bonner, T.I.; Usdin, T.B.

#submission submitted to the EMBL Data Library, October 1995

#accession G02234

#status preliminary; translated from GB/EMBL/DBJ

FEATURE #molecule\_type mRNA  
#residues 1-466 #label BON  
#cross-references EMBL:U39231; NID:G1066050; PID:G1066051

GENETICS GDB:GIPR

#gene #cross-references GDB:335023

#map\_position 19q13.3-19q13.3

SUMMARY #length 466 #molecular-weight 53156 #checksum 265

Query Match 71.4%: Score 55; DB 2; Length 466;  
Best Local Similarity 77.8%: Pred. No. 8.17e+00; Mismatches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 272 IPWIVRYL 280  
111111

Sun Sep 13 10:55:08 1998

US-08-452-843-10.rpt

Page 7

OY 1 PPPIVRL 9

Search completed: Fri Sep 11 13:14:51 1998  
Job time : 30 secs.

---

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
 WIRENET  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Run on: Fri Sep 11 13:13:57 1998: Maspar time 2.43 Seconds  
 Tabular output not generated. 92.731 Million cell updates/sec

Title: >US-08-452-843-10  
 Description: (1-9) from US08452843.pep  
 Perfect Score: 77  
 Sequence: 1 IFFPIVXLY 9

Scoring table: PAM 150  
 Gap 15

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot35  
 1:swiss1

Statistics: Mean 25.814; Variance 31.928; scale 0.809

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID           | Description            | Pred. No. |
|------------|-------|-------------|--------|--------------|------------------------|-----------|
| 1          | 57    | 74.0        | 492    | 1 SP56_BACSU | STAGE V SPOULATION PR  | 7.91e-01  |
| 2          | 56    | 72.7        | 513    | 1 CP12_MESAU | CYTCHROME P450 I2 (E   | 1.26e+00  |
| 3          | 56    | 72.7        | 515    | 1 CP12_CAVPO | CYTCHROME P450 I2 (E   | 1.26e+00  |
| 4          | 56    | 72.7        | 515    | 1 CP12_RABIT | CYTCHROME P450 I2 (E   | 1.26e+00  |
| 5          | 56    | 72.7        | 515    | 1 CP12_HUMAN | CYTCHROME P450 I2 (E   | 1.26e+00  |
| 6          | 56    | 72.7        | 715    | 1 ZP2_CANFA  | ZONA PELLUCIDA SPERM-B | 1.26e+00  |
| 7          | 56    | 72.7        | 716    | 1 ZP2_FELCA  | ZONA PELLUCIDA SPERM-B | 1.26e+00  |
| 8          | 55    | 71.4        | 310    | 1 KITH_HSVTF | THYMIDINE KINASE (EC 2 | 2.01e+00  |
| 9          | 55    | 71.4        | 350    | 1 KITH_HSVTU | THYMIDINE KINASE (EC 2 | 2.01e+00  |
| 10         | 55    | 71.4        | 455    | 1 GIPR_RAT   | GASTRIC INHIBITORY POL | 2.01e+00  |
| 11         | 55    | 71.4        | 462    | 1 GIPR_MESAU | GASTRIC INHIBITORY POL | 2.01e+00  |
| 12         | 55    | 71.4        | 466    | 1 GIPR_HUMAN | GASTRIC INHIBITORY POL | 2.01e+00  |
| 13         | 54    | 70.1        | 513    | 1 CP12_RAT   | CYTCHROME P450 I2 (E   | 3.18e+00  |
| 14         | 54    | 70.1        | 513    | 1 CP12_MOUSE | CYTCHROME P450 I2 (E   | 3.18e+00  |
| 15         | 53    | 68.8        | 666    | 1 ZP2_RABIT  | ZONA PELLUCIDA SPERM-B | 4.98e+00  |
| 16         | 53    | 68.8        | 713    | 1 ZP2_MOUSE  | ZONA PELLUCIDA SPERM-B | 4.98e+00  |
| 17         | 52    | 67.5        | 296    | 1 YHDI_ECOLI | HYPOTHETICAL ADENINE-S | 7.78e+00  |
| 18         | 52    | 67.5        | 1064   | 1 KINI_YEAST | PROTEIN KINASE KINI (E | 1.20e+01  |
| 19         | 51    | 66.2        | 261    | 1 Y046_METUA | HYPOTHETICAL PROTEIN M | 1.20e+01  |
| 20         | 51    | 66.2        | 275    | 1 TR14_ANOGA | TRYPSIN 4 PRECURSOR (E | 1.20e+01  |
| 21         | 51    | 66.2        | 341    | 1 KITH_VZV7  | THYMIDINE KINASE (EC 2 | 1.20e+01  |
| 22         | 51    | 66.2        | 341    | 1 KITH_VZV7  | THYMIDINE KINASE (EC 2 | 1.20e+01  |
| 23         | 51    | 66.2        | 341    | 1 KITH_VZVW  | THYMIDINE KINASE (EC 2 | 1.20e+01  |

|    |    |      |     |              |                         |          |
|----|----|------|-----|--------------|-------------------------|----------|
| 24 | 51 | 66.2 | 341 | 1 KITH_VZV4  | THYMIDINE KINASE (EC 2  | 1.20e+01 |
| 25 | 51 | 66.2 | 341 | 1 KITH_VZV4  | THYMIDINE KINASE (EC 2  | 1.20e+01 |
| 26 | 51 | 66.2 | 350 | 1 E13B_PROPE | GLUCAN ENDO-1,3-BETA-G  | 1.20e+01 |
| 27 | 51 | 66.2 | 366 | 1 GHSR_HUMAN | GROWTH HORMONE SECRETET | 1.20e+01 |
| 28 | 51 | 66.2 | 366 | 1 GHSR_PIG   | GROWTH HORMONE SECRETET | 1.20e+01 |
| 29 | 51 | 66.2 | 380 | 1 FD3E_SOYBN | OMEGA-3 FATTY ACID DES  | 1.20e+01 |
| 30 | 51 | 66.2 | 460 | 1 YS41_CAEEL | HYPOTHETICAL 52.5 KD P  | 1.20e+01 |
| 31 | 51 | 66.2 | 463 | 1 GLPR_RAT   | GLUCAGON-LIKE PEPTIDE   | 1.20e+01 |
| 32 | 51 | 66.2 | 513 | 1 YVPH_BACSU | PUTATIVE L-LACTATE PER  | 1.20e+01 |
| 33 | 51 | 66.2 | 612 | 1 BAL_RAT    | BILE-SALT-ACTIVATED LI  | 1.20e+01 |
| 34 | 51 | 66.2 | 716 | 1 ZP2_PIG    | ZONA PELLUCIDA SPERM-B  | 1.20e+01 |
| 35 | 50 | 64.9 | 220 | 1 UPAS_RAT   | UROKINASE PLASMINOGEN   | 1.85e+01 |
| 36 | 50 | 64.9 | 328 | 1 UPAS_RAT   | UROKINASE PLASMINOGEN   | 1.85e+01 |
| 37 | 50 | 64.9 | 398 | 1 DAP3_HUMAN | DEATH-ASSOCIATED PROTE  | 1.85e+01 |
| 38 | 50 | 64.9 | 453 | 1 TBA1_NEUCR | TUBULIN ALPHA-A CHAIN.  | 1.85e+01 |
| 39 | 50 | 64.9 | 521 | 1 MET3_YEAST | SULFATE ADENYLYLTRANSF  | 1.85e+01 |
| 40 | 50 | 64.9 | 724 | 1 SEC1_YEAST | PROTEIN TRANSPORT PROT  | 1.85e+01 |
| 41 | 50 | 64.9 | 844 | 1 RRPO_IPNVJ | PUTATIVE RNA-DIRECTED   | 1.85e+01 |
| 42 | 50 | 64.9 | 845 | 1 RRPO_IPNVJ | PUTATIVE RNA-DIRECTED   | 1.85e+01 |
| 43 | 49 | 63.6 | 280 | 1 YS33_CAEEL | HYPOTHETICAL 30.5 KD P  | 2.83e+01 |
| 44 | 49 | 63.6 | 285 | 1 SUH2_MOUSE | ALCOHOL G-PROTEIN-COU   | 2.83e+01 |
| 45 | 49 | 63.6 | 401 | 1 Y166_CAEEL | PROBABLE G-PROTEIN-COU  | 2.83e+01 |

## ALIGNMENTS

| RESULT ID | SP56_BACSU   | STANDARD: | PRT: | 492 AA.    |
|-----------|--|-----------|------|------------|
| AC        | P31845;  |           |      |            |
| DT        | 01-JUL-1993 (REL. 26, CREATED)   |           |      |            |
| DT        | 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)                            |           |      |            |
| DT        | 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)                          |           |      |            |
| DE        | STAGE V SPOULATION PROTEIN AF.   |           |      |            |
| GN        | SPOVAF.  |           |      |            |
| OS        | BACILLUS SUBTILIS.   |           |      |            |
| OC        | PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE. |           |      |            |
| RN        | [1]  |           |      |            |
| RP        | SEQUENCE FROM N.A.   |           |      |            |
| RX        | MEDLINE: 85263520.   |           |      |            |
| RL        | FOR P., ERRINGTON J.;  |           |      |            |
| RL        | J. GEN. MICROBIOL. 131:1091-1105(1985).                                |           |      |            |
| RN        | [2]  |           |      |            |
| RP        | SEQUENCE FROM N.A.   |           |      |            |
| RC        | STRAIN-168 / JH642;  |           |      |            |
| RA        | KOBAYASHI Y., MIZUNO M., MASUDA S., TAKEMARU K., HOSONO S.,            |           |      |            |
| RA        | SATO T., TAKEUCHI M.;  |           |      |            |
| RL        | SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.                   |           |      |            |
| RN        | [3]  |           |      |            |
| RP        | SEQUENCE OF 81-492 FROM N.A.   |           |      |            |
| RC        | STRAIN-168 / MAREBURG;   |           |      |            |
| RX        | MEDLINE: 95020538.   |           |      |            |
| RL        | SOROKIN A.V., ZIMSTEIN E., AZEVEDO V., EHRLICH S.D., SERROR P.;        |           |      |            |
| RL        | MOL. MICROBIOL. 10:385-395(1993).                                      |           |      |            |
| RN        | [4]  |           |      |            |
| RP        | SEQUENCE OF 223-492 FROM N.A.  |           |      |            |
| RA        | MEDLINE: 91345841.   |           |      |            |
| RA        | YAMAMOTO J., SHIMIZU M., YAMANE K.;                                    |           |      |            |
| RA        | AGRIC. BIOL. CHEM. 55:1615-1626(1991).                                 |           |      |            |
| CC        | -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).        |           |      |            |
| CC        | -1- SIMILARITY: SOME TO SPORE GERMINATION PROTEIN GERAA AND GERBA.     |           |      |            |
| DR        | EMBL: M15549; -- NOT ANNOTATED_CDS.                                    |           |      |            |
| DR        | EMBL: D84432; G410406; --  |           |      |            |
| DR        | EMBL: L09228; G410115; --  |           |      |            |
| DR        | EMBL: D90189; G216390; --  |           |      |            |
| DR        | PIR: P50430; P50430.   |           |      |            |
| DR        | SUBTILIST: BG10508; SPOVAF.  |           |      |            |
| KV        | SPOULATION: TRANSMEMBRANE.   |           |      |            |
| FT        | TRANSMEM 106 126   |           |      | POTENTIAL. |
| FT        | TRANSMEM 252 272   |           |      | POTENTIAL. |
| FT        | TRANSMEM 296 316   |           |      | POTENTIAL. |
| FT        | TRANSMEM 335 355   |           |      | POTENTIAL. |
| FT        | TRANSMEM 363 383   |           |      | POTENTIAL. |
| FT        | TRANSMEM 387 407   |           |      | POTENTIAL. |

FT CONFLICT 432 432 L -> I (IN REF. 2).  
 FT CONFLICT 467 467 R -> P (IN REF. 4).  
 SQ SEQUENCE 492 AA: 55606 MW: 6CBAAC32 CRC32:

Query Match  
 Best Local Similarity 74.0%; Score 57; DB 1; Length 492;  
 Matches 6; Conservative 1; Pred. No. 7.91e-01; Mismatches 0; Indels 0; Gaps 0;

DB 232 PFPLVRY 238  
 |||:||||  
 OY 2 PFPLVRY 8

RESULT 2  
 ID CP12-MESAU 1 STANDARD; PRT: 513 AA.

AC P24453;  
 DT 01-MAR-1992 (REL. 21, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE CYTOCHROME P450 IA2 (EC 1.14.14.1) (P450-MC4) (METHYLCHOLANTHRENE-  
 INDUCIBLE) (HEPATIC CYTOCHROME P-450MC1).  
 GN CYP1A2.  
 OS MESOCRICETUS AURATUS (GOLDEN HAMSTER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN EUTHERIA; RODENTIA.

RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 91112759.  
 RA LAIT.T.S., CHIANG J.Y.;  
 RL ANCH. BIOCHEM. BIOPHYS. 283:429-439(1990).  
 RN [2].

RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92138673.  
 RA SAGAMI T., OHMACHI T., FUJII H., KIKUCHI H., WATANABE M.;  
 RL J. BIOCHEM. 110:641-647(1991).

CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
 MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
 NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
 OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
 ACIDS, AND XENOBIOTICS.  
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +  
 OXIDIZED FLAVOPROTEIN + H(2)O.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.  
 CC -1- TISSUE SPECIFICITY: FOUND IN LUNG AND LIVER.  
 CC -1- INDUCTION: BY 3-METHYLCHOLANTHRENE.

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 DR EMBL: M63787; G191355; -;  
 DR EMBL: D10252; G220319; -;  
 DR EMBL: D10914; G398133; -;  
 DR PIR: S13885; S13885.

DR PROSITE: PS00086; CYTOCHROME P450; 1.  
 KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME;  
 KM MICROSOME.

FT BINDING 456 456 HEME (BY SIMILARITY).  
 FT CONFLICT 49 49 I -> F (IN REF. 1).  
 FT CONFLICT 52 53 HV -> MC (IN REF. 1).  
 FT CONFLICT 253 254 KN -> GG (IN REF. 1).  
 FT CONFLICT 326 326 L -> W (IN REF. 1).  
 FT CONFLICT 356 356 R -> L (IN REF. 1).  
 FT CONFLICT 485 485 T -> Q (IN REF. 1).  
 SQ SEQUENCE 513 AA: 58082 MW: 40F0041D CRC32;

Query Match  
 Best Local Similarity 72.7%; Score 56; DB 1; Length 513;  
 Matches 6; Conservative 1; Pred. No. 1.26e+00; Mismatches 0; Indels 0; Gaps 0;

DB 238 PFPLVRY 244  
 |||:||||  
 OY 3 PFPLVRY 9

RESULT 3

ID CP12-CAVPO STANDARD; PRT: 515 AA.

AC 064391; 064404;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CYTOCHROME P450 IA2 (EC 1.14.14.1).  
 GN CYP1A2.  
 OS CAVIA PORCELLUS (GUINEA PIG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN EUTHERIA; RODENTIA.

RP SEQUENCE FROM N.A.  
 RC STRAIN-HARTLEY;  
 RA MORI T., ITOH S., OHGITA S., ISHIZAKI K., KAMATARI T.;  
 RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

RP SEQUENCE FROM N.A.  
 RC STRAIN-HARTLEY; TISSUE-LIVER;  
 RA BLACK V.H.;  
 RL [2].

CC -1- FUNCTION (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
 NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
 OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
 ACIDS, AND XENOBIOTICS.  
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +  
 OXIDIZED FLAVOPROTEIN + H(2)O.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 DR EMBL: D50457; G801742; -;  
 DR EMBL: U23501; G790258; -;

DR PROSITE: PS00086; CYTOCHROME P450; 1.  
 KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME;  
 KM MICROSOME.

FT BINDING 458 458 HEME (BY SIMILARITY).  
 FT CONFLICT 7 7 L -> V (IN REF. 2).  
 FT CONFLICT 45 45 P -> A (IN REF. 2).  
 FT CONFLICT 49 49 P -> A (IN REF. 2).  
 FT CONFLICT 149 149 V -> L (IN REF. 2).  
 FT CONFLICT 293 293 E -> Q (IN REF. 2).  
 SQ SEQUENCE 515 AA: 58422 MW: 93E6811E CRC32;

Query Match  
 Best Local Similarity 72.7%; Score 56; DB 1; Length 515;  
 Matches 6; Conservative 1; Pred. No. 1.26e+00; Mismatches 0; Indels 0; Gaps 0;

DB 239 PFPLVRY 245  
 |||:||||  
 OY 3 PFPLVRY 9

RESULT 4  
 ID CP12-RABIT STANDARD; PRT: 515 AA.

AC P00187;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
 DE CYTOCHROME P450 IA2 (EC 1.14.14.1) (ISOZYME 4) (P450-PM4) (TCDD-  
 INDUCIBLE) (BETA-NAPHTHOFLAVONE-INDUCIBLE).  
 GN CYP1A2.  
 OS ORYCTOLAGUS CUNICULUS (RABBIT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN EUTHERIA; LAGOMORPHA.

RP SEQUENCE FROM N.A. (CLONE PLM4-1).  
 RC STRAIN-NEW ZEALAND WHITE;  
 RX MEDLINE: 89052697.  
 RA POMPOON D.;

RL EUR. J. BIOCHEM. 177:285-293(1988).  
 RN [2].  
 RP SEQUENCE FROM N.A. (CLONE PHPAH2/AH3).  
 RX MEDLINE: 88032911.  
 RA KAGAWA N., MIHARA K., SATO R.;



RN J. BIOCHEM. 101:1471-1479(1987).  
 DT [3]  
 RP SEQUENCE.  
 RX MEDLINE: 86140205.  
 RA OZOLS J.;  
 RL J. BIOL. CHEM. 261:3965-3979(1986).  
 RN [4]  
 RP SEQUENCE OF 91-514 FROM N.A.  
 RX MEDLINE: 85270514.  
 RA OKINO S.T., QUATROCHI L.C., BARNES H.J., OSANTO S., GRIFFIN K.J.,  
 RA JOHNSON E.F., TURKEY R.H.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 82:5310-5314(1985).  
 RN [5]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE: 84272618.  
 RA FUJITA V.S., BLACK S.D., TARR G.E., KOOP D.R., COON M.J.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 81:4260-4264(1984).  
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
 CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
 CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
 CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
 CC ACIDS, AND XENOBIOTICS.  
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +  
 CC OXIDIZED FLAVOPROTEIN + H(2)O.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.  
 CC -1- INDUCTION: BY BETA-NAPHTHOFLAVONE AND BY TCDD.  
 DR EMBL: X13853; G1533; -;  
 DR EMBL: M11728; G165579; ALT\_SEQ.  
 DR EMBL: X05686; G1341; -;  
 DR PIR: A00187; O4RBN.  
 DR PIR: A00188; A00188.  
 DR PIR: S02038; S02038.  
 DR PIR: B25143; B25143.  
 DR PIR: B27821; B27821.  
 DR PROSITE: P50086; CYTOCHROME\_P450; 1.  
 KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME;  
 KW MICROSOME.  
 FT INIT. MET 0 0  
 FT BINDING 360 364  
 FT ACT. SITE 402 402  
 FT  
 FT BINDING 457 457  
 FT VARIANT 173 173  
 FT VARIANT 232 232  
 FT VARIANT 298 298  
 FT CONFLICT 21 21  
 FT CONFLICT 69 69  
 FT CONFLICT 91 91  
 FT CONFLICT 120 120  
 FT CONFLICT 171 171  
 FT CONFLICT 193 193  
 FT CONFLICT 207 211  
 FT CONFLICT 246 250  
 FT CONFLICT 288 301  
 FT CONFLICT 353 354  
 FT CONFLICT 357 357  
 FT CONFLICT 358 358  
 FT CONFLICT 461 461  
 FT CONFLICT 493 493  
 SQ SEQUENCE 515 AA; 58202 MW; 552A615C CRC32;  
 Query Match 72.7%; Score 56; DB 1; Length 515;  
 Best Local Similarity 85.7%; Pred. No. 1,26e+00;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

AC P05177;  
 DT 13-AUG-1987 (REL. 05, CREATED)  
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
 DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)  
 DE CYTOCHROME P450 1A2 (EC 1.14.14.1) (P450-P3) (P450-4).  
 GN CYP1A2.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 86312938.  
 RA JAISWAL A.K., NEBERT D.W., GONZALEZ F.J.;  
 RL NUCLEIC ACIDS RES. 14:6773-6774(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 90114205.  
 RA IKERA K., JAISWAL A.K., OWENS R.A., JONES J.E., NEBERT D.W.,  
 RA KIMURA S.;  
 RL MOL. ENDOCRINOL. 3:1399-1408(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 88061719.  
 RA JAISWAL A.K., NEBERT D.W., MCBRIDE O.W., GONZALEZ F.J.;  
 RL J. EXP. PATHOL. 3:1-17(1987).  
 RN [4]  
 RP SEQUENCE OF 295-485 FROM N.A.  
 RX MEDLINE: 86081170.  
 RA QUATROCHI L.C., OKINO S.T., PENDURTHI V.R., TURKEY R.H.;  
 RL DNA 4:395-400(1985).  
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
 CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
 CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
 CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
 CC ACIDS, AND XENOBIOTICS.  
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +  
 CC OXIDIZED FLAVOPROTEIN + H(2)O.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.  
 CC -1- INDUCTION: P450 CAN BE INDUCED TO HIGH LEVELS IN LIVER AND OTHER  
 CC TISSUES BY VARIOUS FOREIGN COMPOUNDS, INCLUDING DRUGS, PESTICIDES,  
 CC AND CARCINOGENS.  
 DR EMBL: Z00036; G30339; -;  
 DR EMBL: M31667; G181382; -;  
 DR EMBL: M31664; G181382; JOINED.  
 DR EMBL: M31665; G181382; JOINED.  
 DR EMBL: M31666; G181382; JOINED.  
 DR EMBL: M12078; G553246; -;  
 DR EMBL: M55053; G181308; -;  
 DR PIR: S07373; O4HU4.  
 DR PIR: S16718; S16718.  
 DR PIR: S22433; S22433.  
 DR MIM: 108330; -;  
 DR MIM: 124060; -;  
 DR PROSITE: P50086; CYTOCHROME\_P450; 1.  
 KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME;  
 KW MICROSOME.  
 FT BINDING 458 458  
 FT CONFLICT 311 311  
 FT CONFLICT 450 451  
 FT CONFLICT 451 451  
 SQ SEQUENCE 515 AA; 58294 MW; C87C0B4C CRC32;  
 Query Match 72.7%; Score 56; DB 1; Length 515;  
 Best Local Similarity 85.7%; Pred. No. 1,26e+00;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 5  
 ID CP12\_HUMAN STANDARD; PRT; 515 AA.

RESULT 6  
 ID ZP2\_CANFA STANDARD; PRT; 715 AA.

```

DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE ZONA PELLUCIDA SPERM-BINDING PROTEIN 2 PRECURSOR (ZONA PELLUCIDA
DE GLYCOPROTEIN ZP2) (ZONA PELLUCIDA PROTEIN A).
CN ZP2 OR ZPA.
OS CANIS FAMILIARIS (DOG).
OC EUARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; CARNIVORA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RX MEDLINE; 95143578.
RA HARRIS J.D., HIBLER D.W., FONTENOT G.K., HSU K.T., YURENICK E.C.,
RA SACC0 A.G.;
RL DNA SEQ. 4:361-393(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RA OKAZAKI Y., ISOJIMA S., SUGIMOTO M.;
RA SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U05779; G458275; -
DR EMBL; D45069; G633050; -
DR PROSITE; PS00682; ZP_DOMAIN: 1.
KW GLYCOPROTEIN; SIGNAL; SULFATATION; SPERM; RECEPTOR; TRANSMEMBRANE;
KW EXTRACELLULAR MATRIX.
FT SIGNAL 1 38 BY SIMILARITY.
FT CHAIN 39 715 ZONA PELLUCIDA SPERM-BINDING PROTEIN 2.
FT DOMAIN 39 684 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 685 705 POTENTIAL.
FT DOMAIN 706 715 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 367 634 ZP.
FT CARBOHYD 87 87 POTENTIAL.
FT CARBOHYD 193 193 POTENTIAL.
FT CARBOHYD 220 220 POTENTIAL.
FT CARBOHYD 266 266 POTENTIAL.
FT CARBOHYD 321 321 POTENTIAL.
FT CONFLICT 15 15 R -> W (IN REF. 2).
FT CONFLICT 292 292 R -> A (IN REF. 2).
FT CONFLICT 328 328 L -> P (IN REF. 2).
FT CONFLICT 599 599 S -> A (IN REF. 2).
SQ SEQUENCE 715 AA; 79938 MW; 6A904635 CRC32;

Query Match 72.7%; Score 56; DB 1; Length 715;
Best Local Similarity 71.4%; Pred. No. 1.26e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 511 YPVRYL 517
QY 3 FPIVRYL 9

RESULT 7
ID ZP2_FELCA STANDARD; PRT; 716 AA.
AC P47984;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ZONA PELLUCIDA SPERM-BINDING PROTEIN 2 PRECURSOR (ZONA PELLUCIDA
DE GLYCOPROTEIN ZP2) (ZONA PELLUCIDA PROTEIN A).
CN ZP2 OR ZPA.
OS FELIS SILVESTRIS CATUS (CAT).
OC EUARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; CARNIVORA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RX MEDLINE; 95143578.
RA HARRIS J.D., HIBLER D.W., FONTENOT G.K., HSU K.T., YURENICK E.C.,
RA SACC0 A.G.;
RL DNA SEQ. 4:361-393(1994).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE-Ovary;
RA OKAZAKI Y., ISOJIMA S., SUGIMOTO M.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: ZP2 FORMS WITH ZP1 AND ZP3 THE ZONA PELLUCIDA, IN
CC WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC ZP2 ACTS AS A SECONDARY SPERM RECEPTOR.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXTRACELLULAR
CC MATRIX.
CC -1- PFM: SULFATED GLYCOPROTEIN WITH O-LINKED OLIGOSACCHARIDES
CC (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS A ZP DOMAIN, WHICH CURRENTLY HAS BEEN FOUND
CC IN ZP2, ZP3, GP2, TGF-3 AND UROMODULIN.
DR EMBL; U05776; G458269; -
DR EMBL; D45067; G633034; -
DR PROSITE; PS00682; ZP_DOMAIN: 1.
KW GLYCOPROTEIN; SIGNAL; SULFATATION; SPERM; RECEPTOR; TRANSMEMBRANE;
KW EXTRACELLULAR MATRIX.
FT SIGNAL 1 38 BY SIMILARITY.
FT CHAIN 39 716 ZONA PELLUCIDA SPERM-BINDING PROTEIN 2.
FT DOMAIN 39 686 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 687 707 POTENTIAL.
FT DOMAIN 708 716 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 369 636 ZP.
FT CARBOHYD 87 87 POTENTIAL.
FT CARBOHYD 96 96 POTENTIAL.
FT CARBOHYD 222 222 POTENTIAL.
FT CARBOHYD 268 268 POTENTIAL.
FT CARBOHYD 531 531 POTENTIAL.
FT CONFLICT 397 397 V -> G (IN REF. 2).
FT CONFLICT 483 483 L -> P (IN REF. 2).
FT CONFLICT 637 637 F -> S (IN REF. 2).
SQ SEQUENCE 716 AA; 80135 MW; F3A1145B CRC32;

Query Match 72.7%; Score 56; DB 1; Length 716;
Best Local Similarity 71.4%; Pred. No. 1.26e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 513 YPVRYL 519
QY 3 FPIVRYL 9

RESULT 8
ID KITH_HSVTF STANDARD; PRT; 310 AA.
AC P13157;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
DE THYMIDINE KINASE (EC 2.7.1.21).
CN TK.
OS TURKEY HERPESVIRUS (STRAIN FC126).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89259069.
RA MARTIN S.L., APARISIO D.I., BANDYOPADHYAY P.K.;
RL J. VIROL. 63:2847-2852(1989).
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE -> ADP + THYMIDINE
CC 5'-PHOSPHATE.
DR EMBL; M26659; G330941; -
DR PIR; A33346; KIBETH.
KW TRANSFERASE; KINASE; DNA SYNTHESIS; ATP-BINDING.
FT NP_BIND 17 24 ATP (PROBABLE).
SQ SEQUENCE 310 AA; 35512 MW; 92745106 CRC32;

Query Match 71.4%; Score 55; DB 1; Length 310;
Best Local Similarity 85.7%; Pred. No. 2.01e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 131 FPIVRYL 137
QY 3 FPIVRYL 9

```

|                       |  |           |                     |                 |  |
|-----------------------|--|-----------|---------------------|-----------------|--|
| FT                    | DOMAIN   | 240       | 251                 |                 | CYTOPLASMIC (POTENTIAL).                 |
| FT                    | TRANSMEM   | 252       | 275                 |                 | 4 (POTENTIAL).                           |
| FT                    | DOMAIN   | 276       | 290                 |                 | EXTRACELLULAR (POTENTIAL).               |
| FT                    | TRANSMEM   | 291       | 316                 |                 | 5 (POTENTIAL).                           |
| FT                    | DOMAIN   | 317       | 338                 |                 | CYTOPLASMIC (POTENTIAL).                 |
| FT                    | TRANSMEM   | 339       | 359                 |                 | 6 (POTENTIAL).                           |
| FT                    | DOMAIN   | 360       | 374                 |                 | EXTRACELLULAR (POTENTIAL).               |
| FT                    | TRANSMEM   | 375       | 395                 |                 | 7 (POTENTIAL).                           |
| FT                    | DOMAIN   | 396       | 455                 |                 | CYTOPLASMIC (POTENTIAL).                 |
| FT                    | CARBOHYD   | 59        | 59                  |                 | POTENTIAL.                               |
| FT                    | CARBOHYD   | 69        | 69                  |                 | POTENTIAL.                               |
| FT                    | CARBOHYD   | 74        | 74                  |                 | POTENTIAL.                               |
| SO                    | SEQUENCE   | 455 AA;   | 52256 MW;           | A435FEF4 CRC32: |  |
| <br>                  |  |           |                     |                 |  |
| Query Match           |  | 71.4%;    | Score 55;           | DB 1;           | Length 455;                              |
| Best Local Similarity |  | 77.8%;    | Pred. No. 2.01e+00; |                 |  |
| Matches               |  | 7;        | Conservative        | 0;              | Mismatches 2; Indels 0; Gaps 0;          |
| <br>                  |  |           |                     |                 |  |
| Db                    | 269 IPFWIVRL 277   |           |                     |                 |  |
|                       |  |           |                     |                 |  |
| QY                    | 1 IPFWIVRL 9   |           |                     |                 |  |
| <br>                  |  |           |                     |                 |  |
| RESULT                | 11   |           |                     |                 |  |
| ID                    | GIPR_MESAU   | STANDARD; | PRT;                | 462 AA.         |  |
| AC                    | P43218;  |           |                     |                 |  |
| DT                    | 01-NOV-1995 (REL. 32, CREATED)   |           |                     |                 |  |
| DT                    | 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  |           |                     |                 |  |
| DT                    | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  |           |                     |                 |  |
| DE                    | GASTRIC INHIBITORY POLYPEPTIDE RECEPTOR PRECURSOR (GIP-R) (GLUCOSE-DEPENDENT INSULINOTROPIC POLYPEPTIDE RECEPTOR).               |           |                     |                 |  |
| GN                    | GIPR.  |           |                     |                 |  |
| OS                    | MESOCRICETUS AURATUS (GOLDEN HAMSTER).   |           |                     |                 |  |
| OC                    | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;   |           |                     |                 |  |
| CC                    | EUTHERIA; RODENTIA.  |           |                     |                 |  |
| RN                    | [1]  |           |                     |                 |  |
| RP                    | SEQUENCE FROM N.A.   |           |                     |                 |  |
| RX                    | MEDLINE; 95110292.   |           |                     |                 |  |
| RA                    | YASUDA K., INAGAKI N., YAMADA Y., KUBOTA A., SEINO S., SEINO Y.;   |           |                     |                 |  |
| RL                    | BIOCHEM. BIOPHYS. RES. COMMUN. 205:1556-1562(1994).  |           |                     |                 |  |
| CC                    | -1- FUNCTION: THIS IS A RECEPTOR FOR GIP. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYL CYCLASE. |           |                     |                 |  |
| CC                    | -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.   |           |                     |                 |  |
| CC                    | -1- TISSUE SPECIFICITY: WIDELY DISTRIBUTED INCLUDING PANCREATIC ISLETS, BRAIN AND VARIOUS PERIPHERAL TISSUES.                    |           |                     |                 |  |
| CC                    | -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  |           |                     |                 |  |
| DR                    | EMBL; D38103; G765087; -.  |           |                     |                 |  |
| DR                    | PROSITE; PS00649; G-PROTEIN_RECE_P2_1; 1.  |           |                     |                 |  |
| DR                    | PROSITE; PS00650; G-PROTEIN_RECE_P2_2; 1.  |           |                     |                 |  |
| KW                    | G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL. POTENTIAL.  |           |                     |                 |  |
| FT                    | SIGNAL   | 1         | 18                  |                 |  |
| FT                    | CHAIN  | 19        | 462                 |                 | GASTRIC INHIBITORY POLYPEPTIDE RECEPTOR. |
| FT                    | DOMAIN   | 19        | 133                 |                 | EXTRACELLULAR (POTENTIAL).               |
| FT                    | TRANSMEM   | 136       | 158                 |                 | 1 (POTENTIAL).                           |
| FT                    | DOMAIN   | 159       | 166                 |                 | CYTOPLASMIC (POTENTIAL).                 |
| FT                    | TRANSMEM   | 167       | 186                 |                 | 2 (POTENTIAL).                           |
| FT                    | DOMAIN   | 187       | 214                 |                 | EXTRACELLULAR (POTENTIAL).               |
| FT                    | TRANSMEM   | 215       | 239                 |                 | 3 (POTENTIAL).                           |
| FT                    | DOMAIN   | 240       | 251                 |                 | CYTOPLASMIC (POTENTIAL).                 |
| FT                    | TRANSMEM   | 252       | 275                 |                 | 4 (POTENTIAL).                           |
| FT                    | DOMAIN   | 276       | 290                 |                 | EXTRACELLULAR (POTENTIAL).               |
| FT                    | TRANSMEM   | 291       | 316                 |                 | 5 (POTENTIAL).                           |
| FT                    | DOMAIN   | 317       | 338                 |                 | CYTOPLASMIC (POTENTIAL).                 |
| FT                    | TRANSMEM   | 339       | 359                 |                 | 6 (POTENTIAL).                           |
| FT                    | DOMAIN   | 360       | 374                 |                 | EXTRACELLULAR (POTENTIAL).               |
| FT                    | TRANSMEM   | 375       | 395                 |                 | 7 (POTENTIAL).                           |
| FT                    | DOMAIN   | 396       | 462                 |                 | CYTOPLASMIC (POTENTIAL).                 |
| FT                    | CARBOHYD   | 59        | 59                  |                 | POTENTIAL.                               |
| FT                    | CARBOHYD   | 74        | 74                  |                 | POTENTIAL.                               |
| SO                    | SEQUENCE   | 462 AA;   | 52918 MW;           | D41J5782 CRC32: |  |

71.4%; Score 55; DB 1; Length 462;

Best Local Similarity 77.8%; Pred. No. 2.01e+00;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 269 IPWIVRYL 277  
1 IPPIVRYL 9

RESULT 12  
ID GIPR HUMAN STANDARD: PRT: 466 AA.  
AC P48546: Q16400: Q14401:  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE GASTRIC INHIBITORY POLYPEPTIDE RECEPTOR PRECURSOR (GIP-R) (GLUCOSE-DEPENDENT INSULINOTROPIC POLYPEPTIDE RECEPTOR).  
GN GIPR.  
OS HOMO SAPIENS (HUMAN).  
OC EURAROTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
OC EUTHERIA: PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA USJIN T.B., GRUBER C., MODI W., BONNER T.I.;  
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96013879.  
RA VOLZ A., GORE R., LANKAT-BUTGEREIT B., FERMAN H.C., BODE H.P.,  
FEBS LETT. 373:23-29(1995).  
RL [3]  
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
RC TISSUE-PANCREAS.  
RX MEDLINE: 96007224.  
RA GREMLICH S., PORRET A., HANI E.H., CHERIF D., VIONNET N., FROGUET P.,  
THORENS B.,  
DIABETES 44:1202-1208(1995).  
RL [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96121393.  
RA YAMADA Y., HAYAMI T., NAKAMURA K., SAISKI P.J., SOMEYA Y.,  
WANG C.Z., SEINO S., SEINO Y.,  
GENOMICS 29:773-776(1995).  
RL [5]  
RP FUNCTION: THIS IS A RECEPTOR FOR GIP. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLYL CYCLASE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
DR EMBL: U39231; G1066051; -  
DR EMBL: S79852; G1184539; -  
DR EMBL: X81832; G1030051; -  
DR EMBL: D49559; G1785516; -  
DR EMBL: D49556; G1785516; JOINED.  
DR EMBL: D49557; G1785516; JOINED.  
DR EMBL: D49558; G1785516; JOINED.  
DR MIM: 137241; -  
DR PROSITE: PS00649; G-PROTEIN\_RECEP\_F2\_1; 1.  
DR PROSITE: PS00650; G-PROTEIN\_RECEP\_F2\_2; 1.  
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;  
CC -1- ALTERNATIVE SPLICING.  
FT CHAIN 1 21 POTENTIAL.  
FT STGNL 1 21 GASTRIC INHIBITORY POLYPEPTIDE RECEPTOR.  
FT DOMAIN 22 466 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 22 138 1 (POTENTIAL).  
FT TRANSSEM 139 161 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 162 169 2 (POTENTIAL).  
FT TRANSSEM 170 189 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 190 217 3 (POTENTIAL).  
FT TRANSSEM 218 242 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 243 254 4 (POTENTIAL).  
FT TRANSSEM 255 278 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 279 293

FT TRANSSEM 294 319 5 (POTENTIAL).  
FT DOMAIN 320 341 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 342 362 6 (POTENTIAL).  
FT DOMAIN 363 377 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 378 398 7 (POTENTIAL).  
FT DOMAIN 399 466 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 62 62 POTENTIAL.  
FT CARBOHYD 77 77 POTENTIAL.  
FT VARSPIC 399 399 V -> VGRDPAAPALRRRGTPAPLSAIVSOV (IN LONG FORM).  
FT CONFLICT 12 12 R -> G (IN REF. 2).  
FT CONFLICT 104 104 G -> R (IN REF. 2).  
FT CONFLICT 117 117 MISSING (IN REF. 3).  
FT CONFLICT 337 337 L -> V (IN REF. 2).  
FT CONFLICT 367 371 GALRF -> APCV (IN REF. 3).  
SQ SEQUENCE 466 AA: 53156 MM; IDC57C17 CRC32;

Query Match 71.4%; Score 55; DB 1; Length 466;  
Best Local Similarity 77.8%; Pred. No. 2.01e+00;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 272 IPWIVRYL 280  
1 IPPIVRYL 9

RESULT 13  
ID CP12\_RAT STANDARD: PRT: 513 AA.  
AC P04799;  
DT 13-AUG-1987 (REL. 05, CREATED)  
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
DE CYTOCHROME P450 1A2 (EC 1.14.14.1) (P450-D) (METHYLCHOLANTHRENE-INDUCIBLE) (P-448).  
GN CYP1A-2.  
OS RATIUS NORVEGICUS (RAT).  
OC EURAROTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
OC EUTHERIA: RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85182625.  
RA SOGAWA K., GOTOH O., KAWAJIRI K., HARADA T., FUJII-KURIYAMA Y.,  
J. BIOL. CHEM. 260:5026-5032(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 84170359.  
RA KAWAJIRI K., GOTOH O., SOGAWA K., TAGASHIRA Y., MURAMATSU M.,  
FUJII-KURIYAMA Y.;  
PROC. NATL. ACAD. SCI. U.S.A. 81:1649-1653(1984).  
RL [3]  
RP FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.  
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) - ROH + OXIDIZED FLAVOPROTEIN + H(2)O.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.  
CC -1- INDUCTION: BY METHYLCHOLANTHRENE.  
DR EMBL: K03241; G203832; -  
DR EMBL: K02422; G203763; -  
DR PIR: A22562; A22562.  
DR PIR: A20963; A20963.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME; MICROsome.  
FT BINDING 456 456 HEME.  
FT CONFLICT 137 137 H -> R (IN REF. 2).  
FT CONFLICT 262 262 F -> S (IN REF. 2).  
FT CONFLICT 403 403 R -> C (IN REF. 2).  
SQ SEQUENCE 513 AA: 58293 MM; B560C869 CRC32;

Query Match 70.1%; Score 54; DB 1; Length 513;  
Best Local Similarity 71.4%; Pred. No. 3.18e+00;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 238 FPLRYL 244  
 ||:||||  
 OY 3 FPIVRYL 9

RESULT 14  
 ID CP12\_MOUSE STANDARD; PRT: 513 AA.  
 AC P00186;  
 DT 21-JUL-1986 (REL. 01, LAST CREATED)  
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CYTOCHROME P450 1A2 (EC 1.14.14.1) (P450-P3 AND P450-P2)  
 DE (METHYLCHOLANTHRENE-INDUCIBLE).  
 GN CYP1A2 OR CYP1A2-2.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 [1]  
 RP SEQUENCE FROM N.A. (P3).  
 RC STRAIN-C57BL/6N;  
 RX MEDLINE: 84289486.  
 RA KIMURA S., GONZALEZ F.J., NEBERT D.W.;  
 RL J. BIOL. CHEM. 259:10705-10713(1984).  
 RN [12]  
 RP SEQUENCE FROM N.A. (P3).  
 RC STRAIN-C57BL/6N;  
 RX MEDLINE: 84169582.  
 RA KIMURA S., GONZALEZ F.J., NEBERT D.W.;  
 RL NUCLEIC ACIDS RES. 12:2917-2928(1984).  
 RN [13]  
 RP SEQUENCE FROM N.A. (P3).  
 RX MEDLINE: 85182627.  
 RA GONZALEZ F.J., KIMURA S., NEBERT D.W.;  
 RL J. BIOL. CHEM. 260:5040-5049(1985).  
 RN [14]  
 RP ERRATUM.  
 RA GONZALEZ F.J., KIMURA S., NEBERT D.W.;  
 RL J. BIOL. CHEM. 260:11884-11889(1985).  
 RN [15]  
 RP SEQUENCE FROM N.A. (P3).  
 RX MEDLINE: 85028449.  
 RA GONZALEZ F.J., MACKENZIE P.I., KIMURA S., NEBERT D.W.;  
 RL GENE 29:281-292(1984).  
 RN [16]  
 RP SEQUENCE FROM N.A. (P2).  
 RC STRAIN-DBA/2N; TISSUE=LIVER;  
 RX MEDLINE: 86312932.  
 RA KIMURA S., NEBERT D.W.;  
 RL NUCLEIC ACIDS RES. 14:6765-6766(1986).  
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.  
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH + OXIDIZED FLAVOPROTEIN + H(2)O.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.  
 DR EMBL: X01682; G50611;  
 DR EMBL: X00479; G50603;  
 DR EMBL: X04283; G50628;  
 DR EMBL: K02589; G309205; ALT\_SEQ.  
 DR EMBL: M10022; G367139;  
 DR PIR: A00186; O4MSM3.  
 DR PIR: B23923; B23923.  
 DR MGD: MGI:88589; CYP1A2.  
 DR PROSITE: PS00086; CYTOCHROME\_P450\_1.  
 KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME; MICROSOME.  
 FT BINDING 456 456 HEME.  
 FT VARIANT 384 384 I -> M (IN P2).

SO SEQUENCE 513 AA; 58184 MW; 98334C3C CRC32;

Query Match  
 Best Local Similarity 71.4%; Score 54; DB 1; Length 513;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 238 FPLRYL 244  
 ||:||||  
 OY 3 FPIVRYL 9

RESULT 15  
 ID ZP2-RABIT STANDARD; PRT: 666 AA.  
 AC P48829;  
 DT 01-FEB-1996 (REL. 33, LAST CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE ZONA PELLUCIDA SPERM-BINDING PROTEIN 2 (ZONA PELLUCIDA GLYCOPROTEIN ZP2) (ZONA PELLUCIDA PROTEIN A) (75 KD ZONA PELLUCIDA PROTEIN)  
 DE (FRAGMENT).  
 GN ZP2 OR ZPA.  
 OS ORCTOLAGUS CUNICULUS (RABBIT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; LAGOMORPHA.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NEW ZEALAND WHITE; TISSUE=OVARY;  
 RX MEDLINE: 93286072.  
 RA LEE V.H., SCHMOEBEL E.D., PRASAD S.V., CHEUNG P., TIMMONS T.M., COOK R.G., DONBAR B.S.;  
 RL J. BIOL. CHEM. 268:12412-12417(1993).  
 CC -1- FUNCTION: ZP2 FORMS WITH ZP1 AND ZP3 THE ZONA PELLUCIDA. IN WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1. ZP2 ACTS AS A SECONDARY SPERM RECEPTOR (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXTRACELLULAR MATRIX.  
 CC -1- PTM: PROTEOLYTICALLY CLEAVED AFTER FERTILIZATION, AND THIS MODIFICATION, ALONG WITH PRESUMED CHANGES IN ZP3 MAY PLAY AN IMPORTANT ROLE IN THE POSTFERTILIZATION BLOCK TO POLYSPERMY.  
 CC -1- PTM: SULFATED GLYCOPROTEIN WITH O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS A ZP DOMAIN, WHICH CURRENTLY HAS BEEN FOUND IN ZP2, ZP3, GP2, TGF-3 AND UROMODULIN.  
 DR EMBL: L12167; G165818;  
 DR PROSITE: PS00682; ZP\_DOMAIN; 1.  
 KW GLYCOPROTEIN; SIGNAL; SULFATATION; SPERM; RECEPTOR; TRANSMEMBRANE; EXTRACELLULAR MATRIX.  
 FT NON\_TER 1 1  
 FT CHAIN 1 1  
 FT DOMAIN <1 666 ZONA PELLUCIDA SPERM-BINDING PROTEIN 2.  
 FT TRANSMEM 637 636 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 657 666 POTENTIAL.  
 FT CARBOHYD 318 585 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 38 38 ZP.  
 FT CARBOHYD 73 73 POTENTIAL.  
 FT CARBOHYD 126 126 POTENTIAL.  
 FT CARBOHYD 171 171 POTENTIAL.  
 FT CARBOHYD 217 217 POTENTIAL.  
 FT CARBOHYD 241 241 POTENTIAL.  
 FT CARBOHYD 613 613 POTENTIAL.  
 SQ SEQUENCE 666 AA; 73644 MW; E8332457 CRC32;

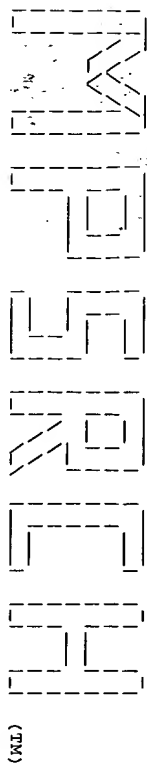
Query Match  
 Best Local Similarity 71.4%; Score 53; DB 1; Length 666;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 462 YPIVRYL 468  
 ||:||||  
 OY 3 FPIVRYL 9

Search completed: Fri Sep 11 13:14:03 1998  
 Job time : 6 secs.

Sun Sep 13 10:55:09 1998

US-08-452-843-10.rsp



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Run on: Fri Sep 11 13:15:08 1998; Maspar time 3.59 seconds  
105.572 Million cell updates/sec

Title: >US-08-452-843-10  
Description: (1-9) from US08452843.pep  
Perfect Score: 77  
Sequence: 1 IFFPIRYL 9

Scoring table: PAM 150  
Gap 15  
Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: 1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mmc 6:sp\_orzanelle 7:sp\_phase 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp\_vertebrate  
13:sp\_unclassified

Statistics: Mean 24.680; Variance 34.214; scale 0.721

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description             | Pred. No. |
|------------|-------|-------------|--------|----|--------|-------------------------|-----------|
| 1          | 59    | 76.6        | 1645   | 3  | Q27448 | GLUTAMINE-DEPENDENT CA  | 8.67e-01  |
| 2          | 56    | 72.7        | 516    | 2  | Q16754 | CYTOKROME P-450 4 (EC   | 3.24e+00  |
| 3          | 56    | 72.7        | 516    | 4  | Q29526 | LIVER CYTOKROME P450    | 3.24e+00  |
| 4          | 55    | 71.4        | 145    | 6  | P92553 | ORF145B.                | 4.99e+00  |
| 5          | 55    | 71.4        | 212    | 9  | P76249 | SIMILAR TO.             | 4.99e+00  |
| 6          | 55    | 71.4        | 333    | 9  | Q34616 | ORF.                    | 4.99e+00  |
| 7          | 55    | 71.4        | 1025   | 10 | Q10836 | THYROTROPIN-RELEASING   | 4.99e+00  |
| 8          | 54    | 70.1        | 341    | 11 | P80988 | THYROIDINE KINASE (EC 2 | 7.63e+00  |
| 9          | 54    | 70.1        | 367    | 10 | Q64588 | CYTOKROME P450 (EC 1.   | 7.63e+00  |
| 10         | 54    | 70.1        | 482    | 3  | Q09998 | POTATIVE 55.5 KD ZINC   | 7.63e+00  |
| 11         | 53    | 68.8        | 189    | 3  | Q17467 | B0284.3.                | 1.16e+01  |
| 12         | 53    | 68.8        | 588    | 3  | P92010 | R10D12.8.               | 1.16e+01  |
| 13         | 52    | 67.5        | 112    | 11 | Q84619 | GENOME, PARTIAL SEQUEN  | 1.76e+01  |
| 14         | 52    | 67.5        | 291    | 8  | P93740 | HYPOTHETICAL 32.4 KD P  | 1.76e+01  |
| 15         | 52    | 67.5        | 1932   | 3  | Q01483 | COSMID C06A5.           | 1.76e+01  |
| 16         | 51    | 66.2        | 210    | 3  | Q18053 | CODED FOR BY C. ELEGAN  | 2.64e+01  |
| 17         | 51    | 66.2        | 231    | 11 | P80985 | THYROIDINE KINASE (EC 2 | 2.64e+01  |
| 18         | 51    | 66.2        | 242    | 9  | Q58987 | PHOSPHORIBOSYLAMINOIMI  | 2.64e+01  |
| 19         | 51    | 66.2        | 295    | 3  | Q93291 | HYPOTHETICAL PROTEIN C  | 2.64e+01  |
| 20         | 51    | 66.2        | 328    | 11 | Q90020 | THYROIDINE KINASE (EC 2 | 2.64e+01  |

|    |    |      |      |    |        |                         |          |
|----|----|------|------|----|--------|-------------------------|----------|
| 21 | 51 | 66.2 | 328  | 11 | Q96697 | THYROIDINE KINASE.      | 2.64e+01 |
| 22 | 51 | 66.2 | 341  | 11 | P80990 | THYROIDINE KINASE (EC 2 | 2.64e+01 |
| 23 | 51 | 66.2 | 341  | 11 | P80991 | THYROIDINE KINASE (EC 2 | 2.64e+01 |
| 24 | 51 | 66.2 | 341  | 11 | P80993 | THYROIDINE KINASE (EC 2 | 2.64e+01 |
| 25 | 51 | 66.2 | 364  | 10 | O08725 | GROWTH HORMONE SECRET   | 2.64e+01 |
| 26 | 51 | 66.2 | 485  | 3  | Q16873 | C13A2.5 PROTEIN.        | 2.64e+01 |
| 27 | 51 | 66.2 | 489  | 10 | Q35659 | GLUCAGON-LIKE PEPTIDE   | 2.64e+01 |
| 28 | 51 | 66.2 | 496  | 3  | Q17676 | C9A1.5.                 | 2.64e+01 |
| 29 | 51 | 66.2 | 503  | 3  | Q23469 | HYPOTHETICAL 58.3 KD P  | 2.64e+01 |
| 30 | 51 | 66.2 | 592  | 10 | Q64571 | STEROL ESTERASE (EC 3.  | 2.64e+01 |
| 31 | 50 | 64.9 | 227  | 9  | P72941 | BIOPOLYMER TRANSPORT    | 3.96e+01 |
| 32 | 50 | 64.9 | 328  | 10 | Q35771 | UNOXYGEN RECEPTOR.      | 3.96e+01 |
| 33 | 50 | 64.9 | 382  | 2  | Q13044 | IONIZING RADIATION RES  | 3.96e+01 |
| 34 | 50 | 64.9 | 488  | 9  | P71364 | GUCCONATE PERMEASE (GN  | 3.96e+01 |
| 35 | 50 | 64.9 | 546  | 3  | Q21733 | R05D11.1.               | 3.96e+01 |
| 36 | 50 | 64.9 | 555  | 8  | Q04693 | GLOSSY1 HOMOLOG (FRAGM  | 3.96e+01 |
| 37 | 50 | 64.9 | 698  | 8  | P95865 | ORF C06018.             | 3.96e+01 |
| 38 | 50 | 64.9 | 845  | 11 | Q82732 | RNA-DEPENDENT RNA POLY  | 5.89e+01 |
| 39 | 49 | 63.6 | 154  | 9  | Q27522 | CONSERVED PROTEIN       | 5.89e+01 |
| 40 | 49 | 63.6 | 177  | 9  | P95521 | RIEKE IRON-SULFUR PRO   | 5.89e+01 |
| 41 | 49 | 63.6 | 221  | 2  | Q13242 | SPLITING FACTOR, ARGIN  | 5.89e+01 |
| 42 | 49 | 63.6 | 238  | 3  | P91540 | COSMID ZC204.           | 5.89e+01 |
| 43 | 49 | 63.6 | 528  | 1  | Q13345 | OXIDOREDUCTASE.         | 5.89e+01 |
| 44 | 49 | 63.6 | 651  | 3  | P91538 | COSMID ZC204.           | 5.89e+01 |
| 45 | 49 | 63.6 | 1109 | 3  | Q23356 | ZC504.4.                | 5.89e+01 |

## ALIGNMENTS

| RESULT | ID     | Query Match | Score | DB | Length | Matches | Indels | Gaps |
|--------|--------|-------------|-------|----|--------|---------|--------|------|
| 1      | Q27448 | 76.6%       | 59    | 3  | 1645   | 77.8%   | 1      | 0    |
| 2      | Q16754 | 72.7%       | 56    | 2  | 516    | 77.8%   | 1      | 0    |
| 3      | Q29526 | 72.7%       | 56    | 4  | 516    | 77.8%   | 1      | 0    |
| 4      | P92553 | 71.4%       | 55    | 6  | 145    | 77.8%   | 1      | 0    |
| 5      | P76249 | 71.4%       | 55    | 9  | 212    | 77.8%   | 1      | 0    |
| 6      | Q34616 | 71.4%       | 55    | 9  | 333    | 77.8%   | 1      | 0    |
| 7      | Q10836 | 71.4%       | 55    | 10 | 1025   | 77.8%   | 1      | 0    |
| 8      | P80988 | 70.1%       | 54    | 11 | 341    | 77.8%   | 1      | 0    |
| 9      | Q64588 | 70.1%       | 54    | 10 | 367    | 77.8%   | 1      | 0    |
| 10     | Q09998 | 70.1%       | 54    | 3  | 482    | 77.8%   | 1      | 0    |
| 11     | Q17467 | 68.8%       | 53    | 3  | 189    | 77.8%   | 1      | 0    |
| 12     | P92010 | 68.8%       | 53    | 3  | 588    | 77.8%   | 1      | 0    |
| 13     | Q84619 | 67.5%       | 52    | 11 | 112    | 77.8%   | 1      | 0    |
| 14     | P93740 | 67.5%       | 52    | 8  | 291    | 77.8%   | 1      | 0    |
| 15     | Q01483 | 67.5%       | 52    | 3  | 1932   | 77.8%   | 1      | 0    |
| 16     | Q18053 | 66.2%       | 51    | 3  | 210    | 77.8%   | 1      | 0    |
| 17     | P80985 | 66.2%       | 51    | 11 | 231    | 77.8%   | 1      | 0    |
| 18     | Q58987 | 66.2%       | 51    | 9  | 242    | 77.8%   | 1      | 0    |
| 19     | Q93291 | 66.2%       | 51    | 3  | 295    | 77.8%   | 1      | 0    |
| 20     | Q90020 | 66.2%       | 51    | 11 | 328    | 77.8%   | 1      | 0    |

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 86313652.  
 RA QUATROCHI L.C., PENDURTHI U.R., OKINO S.T., POTENZA C., TURKEY R.H.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:6731-6735(1986).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 DR EMBL: L00389; G181317; JOINED.  
 DR EMBL: L00384; G181317; JOINED.  
 DR EMBL: L00385; G181317; JOINED.  
 DR EMBL: L00386; G181317; JOINED.  
 DR EMBL: L00387; G181317; JOINED.  
 DR EMBL: L00388; G181317; JOINED.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME.  
 FT BINDING 458 458 HEME (BY SIMILARITY).  
 SQ SEQUENCE 516 AA; 58279 MW; 96B7DA19 CRC32;  
 Query Match  
 Best Local Similarity 72.7%; Score 56; DB 2; Length 516;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 239 FPIRLYL 245  
 |||:||||  
 3 FPIVRYL 9  
 RESULT 3  
 ID Q29526 PRELIMINARY; PRT; 516 AA.  
 AC Q29526;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE LAYER CYTOCHROME P450 LM4 (EC. 1.14.14.1).  
 OS ORCOTOLAGUS CUNICULUS (RABBIT).  
 CC EURAROTIA; METACORA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EURAROTIA; METACORA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89052697.  
 RA POMRON D.;  
 RL EUR. J. BIOCHEM. 177:285-293(1988).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 DR EMBL: M36538; G165591;  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME.  
 FT BINDING 458 458 HEME (BY SIMILARITY).  
 SQ SEQUENCE 516 AA; 58212 MW; A31D999F CRC32;  
 Query Match  
 Best Local Similarity 72.7%; Score 56; DB 4; Length 516;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 239 FPIRLYL 245  
 |||:||||  
 3 FPIVRYL 9  
 RESULT 4  
 ID P92553 PRELIMINARY; PRT; 145 AA.  
 AC P92553;  
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
 DE 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)  
 DE ORP145B  
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OG MITOCHONDRION.  
 CC EURAROTIA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;  
 OC CAPPALES; CRUCIFERAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA UNSELD M., MARLENFELD J.R., BRANDT P., BRENNICKE A.;  
 RL NATURE GENET. 0:0-0(0).  
 DR EMBL: Y08502; E283485;

KW MITOCHONDRION.  
 SQ SEQUENCE 145 AA; 17046 MW; 47C0CF7F CRC32;  
 Query Match  
 Best Local Similarity 71.4%; Score 55; DB 6; Length 145;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 132 IPEFSRYL 140  
 |||:||||  
 1 IPEIVRYL 9  
 RESULT 5  
 ID P76249 PRELIMINARY; PRT; 212 AA.  
 AC P76249; 007971; 007969;  
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
 DE SIMILAR TO.  
 OS ESCHERICHIA COLI.  
 CC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K-12;  
 RA ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,  
 RA KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M., MAKINO K., MIKI T.,  
 RA MIZOBUCHI K., MORI H., MORI T., MOTOMURA K., NAKADE S., NAKAMURA Y.,  
 RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPET G., SEKI Y.,  
 RA SIVASUNDARAM S., TAGAMI H., TAKEDA J., TAKEMOTO K., WADA C.,  
 RA YAMAMOTO Y., HORIUCHI T.;  
 RL DNA RES. 3:379-392(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONZO A., HORIUCHI T.,  
 RA IEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K., KASAI H.,  
 RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,  
 RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,  
 RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPET G., SEKI Y.,  
 RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.;  
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AE000274; G1788099;  
 DR EMBL: D90824; G1736431;  
 DR EMBL: D90823; G1736421;  
 SQ SEQUENCE 212 AA; 23200 MW; E748AC96 CRC32;  
 Query Match  
 Best Local Similarity 71.4%; Score 55; DB 9; Length 212;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Db 74 IPEIVRYL 82  
 |||:||||  
 1 IPEIVRYL 9  
 RESULT 6  
 ID O34616 PRELIMINARY; PRT; 323 AA.  
 AC O34616;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE OREB.  
 GN YVAX.  
 OS BACILLUS SUBTILIS.  
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.



[illegible]

RESULT 9  
ID 064588 PRELIMINARY; PRT: 367 AA.  
AC 064588  
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
DE 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
OC CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85054751.  
RA YABUSAKI Y., MURAKAMI H., NAKAMURA K., NOMURA N., SHIMIZU M., OEDA K.,  
RA OKAWA H.;  
RL J. BIOCHEM. 96:793-804(1984)  
CC EMBL: X01031; 6809075;  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME.  
FT NON TER 1  
FT BINDING 310 310 HEME (BY SIMILARITY).  
SQ SEQUENCE 367 AA; 42235 MW; F84516FC CRC32;  
Query Match 70.1%; Score 54; DB 10; Length 367;  
Best Local Similarity 71.4%; Pred. No. 7.63e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 92 FPIVRYL 98  
QY 3 FPIVRYL 9  
RESULT 10  
ID 009998 PRELIMINARY; PRT: 492 AA.  
AC 009998  
DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
DT 01-MAY-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMELREL. 05, LAST ANNOTATION UPDATE)  
DE PUTATIVE 55.5 KD ZINC FINGER PROTEIN R144.3 IN CHROMOSOME III.  
GN R144.3  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA FAVELLO T.;  
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC EMBL: U23515; 6746495;  
DR WORMPEP: R144.3; CE02033.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 2.  
KW HYPOTHETICAL PROTEIN; ZINC-FINGER; DNA-BINDING; METAL-BINDING;  
KW NUCLEAR-PROTEIN.  
FT DOMAIN 230 287 ZINC-FINGERS.  
FT ZN\_FING 228 253 C2H2-TYPE.  
FT ZN\_FING 264 287 C2H2-TYPE.  
SQ SEQUENCE 492 AA; 55479 MW; D62CA443 CRC32;  
Query Match 70.1%; Score 54; DB 3; Length 492;  
Best Local Similarity 85.7%; Pred. No. 7.63e+00;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 109 FQIVRYL 115  
QY 3 FPIVRYL 9  
RESULT 11  
ID 017467 PRELIMINARY; PRT: 189 AA.  
AC 017467  
DT 01-NOV-1996 (TREMELREL. 01, CREATED)

DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMELREL. 01, LAST ANNOTATION UPDATE)  
DE B0284.3  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RL SULSTON J.;  
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA LIGHTNING J., LLOYD C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 368:32-38(1994).  
DR EMBL: Z30973; G463202;  
SQ SEQUENCE 189 AA; 21347 MW; C31E7906 CRC32;  
Query Match 68.8%; Score 53; DB 3; Length 189;  
Best Local Similarity 75.0%; Pred. No. 1.16e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 23 FPIVRYL 30  
QY 2 FPIVRYL 9  
RESULT 12  
ID P92010 PRELIMINARY; PRT: 588 AA.  
AC P92010  
DT 01-MAY-1997 (TREMELREL. 03, CREATED)  
DT 01-MAY-1997 (TREMELREL. 03, LAST SEQUENCE UPDATE)  
DT 01-MAY-1997 (TREMELREL. 03, LAST ANNOTATION UPDATE)  
DE R10D12.8  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA PERCY C.;  
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA LIGHTNING J., LLOYD C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 368:32-38(1994).  
DR EMBL: 281109; E308118;  
SQ SEQUENCE 588 AA; 69273 MW; 697E93DF CRC32;  
Query Match 68.8%; Score 53; DB 3; Length 588;  
Best Local Similarity 55.6%; Pred. No. 1.16e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 108 VSEPIVRYL 116  
QY 1 IPEPIVRYL 9

```

RESULT 13
ID .084619; PRELIMINARY; PRT; 112 AA.
AC 084619;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE GENOME, PARTIAL SEQUENCE.
GN A303L.
OS PARAMECIUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PHYCODNAVIRIDAE.
RA [1].
RP SEQUENCE FROM N.A.
RX MEDLINE: 95133167.
RA LU Z., LI Y., ZHANG Y., KUTISH G.F., ROCK D.L., VAN ETEN J.L.;
RL VIROLOGY 206:339-352(1995).
DR EMBL: U42580; G1181466; -
SQ SEQUENCE 112 AA; 13416 MW; 5C07006C CRC32;

Query Match 67.5%; Score 52; DB 11; Length 112;
Best Local Similarity 55.6%; Pred. No. 1.76e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 20 VPSIIRNL 28
QY 1 IPPIVRYL 9

RESULT 14
ID P93740; PRELIMINARY; PRT; 291 AA.
AC P93740;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE HYPOTHEICAL 32.4 KD PROTEIN.
GN T06D20.4.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONAE;
OC CAPRIFALES; CRUCIFERAE.
RA [1].
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA ROUNSLEY S.D., KETCHUM K.A., LIN X., PHILLIPS C.A., BRANDON R.C.,
RA FUHRMANN J.L., KERLAVAGE A.R., ADAMS M.D., SOMERVILLE C.R.,
RA VENTER J.C.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U90439; G1871177; -
KW HYPOTHEICAL PROTEIN.
SQ SEQUENCE 291 AA; 32374 MW; C8365B16 CRC32;

Query Match 67.5%; Score 52; DB 8; Length 291;
Best Local Similarity 62.5%; Pred. No. 1.76e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 254 PEPFTRFL 261
QY 2 PPIVRYL 9

RESULT 15
ID 001483; PRELIMINARY; PRT; 1932 AA.
AC 001483;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)
DE COSMID C06A5.
GN C06A5.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RA [1].
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE: 94150718.

```

```

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,
RA COULSON A., CRAYTON M., DEAR S., DU Z., DURBIN R., FAYELLO A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIKEN L., ROOPRA A.,
RA SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONNHAMMER E.,
RA STADEN R., SUSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
RA [2].
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA DAVIDSON S., WOHLDMANN P.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3].
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U97193; G1943786; -
SQ SEQUENCE 1932 AA; 219514 MW; 1A402583 CRC32;

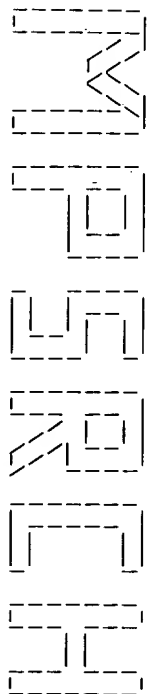
Query Match 67.5%; Score 52; DB 3; Length 1932;
Best Local Similarity 71.4%; Pred. No. 1.76e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 733 YPIRYL 739
QY 3 FPIVRYL 9

Search completed: Fri Sep 11 13:15:43 1998
Job time : 35 secs.

```

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

MPearch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:09:35 1998; Maspar time 2.56 Seconds  
56.822 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-9  
Description: (1-9) from US08452843.pep  
Perfect Score: 70  
Sequence: 1 IYPYIVRS L 9

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

a-geneseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 17.010; Variance 46.169; scale 0.368

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length  | DB ID  | Description            | Pred. No. |
|------------|-------|-------------|---------|--------|------------------------|-----------|
| 1          | 70    | 100.0       | 9 18    | R89370 | Cw6 consensus peptide  | 1.47e-01  |
| 2          | 65    | 92.9        | 9 18    | R89369 | Cw6 consensus peptide  | 6.41e-01  |
| 3          | 57    | 81.4        | 9 18    | R89371 | Cw6 consensus peptide  | 6.33e-00  |
| 4          | 52    | 74.3        | 133 25  | W28511 | Product of clone L105  | 2.53e-01  |
| 5          | 50    | 71.4        | 356 2   | P70388 | D-amino acid oxidase   | 4.36e-01  |
| 6          | 50    | 71.4        | 356 3   | R04066 | T. variabilis D-amino  | 4.36e-01  |
| 7          | 50    | 71.4        | 1676 15 | R77604 | Pro-c5 polypeptide     | 4.36e-01  |
| 8          | 47    | 67.1        | 311 25  | W25084 | Haemophilus influenza  | 9.72e-01  |
| 9          | 47    | 67.1        | 1082 13 | R65017 | PR2 retinoblastoma t   | 9.72e-01  |
| 10         | 47    | 67.1        | 1230 23 | W17785 | Potato tuber solute    | 9.72e-01  |
| 11         | 46    | 65.7        | 225 21  | W15422 | G protein conjugative  | 1.26e-02  |
| 12         | 46    | 65.7        | 225 26  | W35832 | Plasmid pUD-B15 human  | 1.26e-02  |
| 13         | 46    | 65.7        | 524 27  | W35346 | Arabidopsis thaliana   | 1.26e-02  |
| 14         | 46    | 65.7        | 1330 3  | R13444 | Swine herpes virus-1   | 1.26e-02  |
| 15         | 45    | 64.3        | 210 3   | R13499 | P. denitrificans COB H | 1.64e-02  |
| 16         | 45    | 64.3        | 713 11  | R60101 | Canine zona pellucida  | 1.64e-02  |
| 17         | 45    | 64.3        | 716 12  | R60532 | Feline zona pellucida  | 1.64e-02  |
| 18         | 45    | 64.3        | 1592 27 | W34623 | Human C3 protein muta  | 1.64e-02  |

|    |    |      |         |        |                       |          |
|----|----|------|---------|--------|-----------------------|----------|
| 19 | 45 | 64.3 | 1635 27 | W34624 | Human C3 protein muta | 1.64e-02 |
| 20 | 45 | 64.3 | 1657 27 | W34629 | Human C3 protein muta | 1.64e-02 |
| 21 | 45 | 64.3 | 1661 27 | W34625 | Human C3 protein muta | 1.64e-02 |
| 22 | 45 | 64.3 | 1663 27 | W34620 | Human C3 protein muta | 1.64e-02 |
| 23 | 45 | 64.3 | 1663 27 | W34614 | Human C3 protein muta | 1.64e-02 |
| 24 | 45 | 64.3 | 1663 27 | W34608 | Human C3 protein muta | 1.64e-02 |
| 25 | 45 | 64.3 | 1663 27 | W34607 | Human C3 protein muta | 1.64e-02 |
| 26 | 45 | 64.3 | 1663 27 | W34618 | Human C3 protein muta | 1.64e-02 |
| 27 | 45 | 64.3 | 1663 27 | W40989 | Human C3 protein muta | 1.64e-02 |
| 28 | 45 | 64.3 | 1663 27 | W34606 | Wild type human C3 pr | 1.64e-02 |
| 29 | 45 | 64.3 | 1663 27 | W34621 | Human C3 protein muta | 1.64e-02 |
| 30 | 45 | 64.3 | 1663 27 | W34630 | Human C3 protein muta | 1.64e-02 |
| 31 | 45 | 64.3 | 1663 27 | W34619 | Human C3 protein muta | 1.64e-02 |
| 32 | 45 | 64.3 | 1663 27 | W40988 | Human C3 protein muta | 1.64e-02 |
| 33 | 45 | 64.3 | 1663 27 | W34628 | Human C3 protein muta | 1.64e-02 |
| 34 | 45 | 64.3 | 1663 27 | W34627 | Human C3 protein muta | 1.64e-02 |
| 35 | 45 | 64.3 | 1663 27 | W34616 | Human C3 protein muta | 1.64e-02 |
| 36 | 45 | 64.3 | 1663 27 | W40990 | Human C3 protein muta | 1.64e-02 |
| 37 | 45 | 64.3 | 1663 27 | W34609 | Human C3 protein muta | 1.64e-02 |
| 38 | 45 | 64.3 | 1663 27 | W34610 | Human C3 protein muta | 1.64e-02 |
| 39 | 45 | 64.3 | 1663 27 | W34611 | Human C3 protein muta | 1.64e-02 |
| 40 | 45 | 64.3 | 1663 27 | W34612 | Human C3 protein muta | 1.64e-02 |
| 41 | 45 | 64.3 | 1663 27 | W34615 | Human C3 protein muta | 1.64e-02 |
| 42 | 45 | 64.3 | 1663 27 | W34617 | Human C3 protein muta | 1.64e-02 |
| 43 | 45 | 64.3 | 1663 27 | W34613 | Human C3 protein muta | 1.64e-02 |
| 44 | 45 | 64.3 | 1667 27 | W34626 | Human C3 protein muta | 1.64e-02 |
| 45 | 45 | 64.3 | 1667 27 | W34631 | Human C3 protein muta | 1.64e-02 |

## ALIGNMENTS

RESULT 1  
ID R89370 standard; peptide: 9 AA.  
AC R89370;  
DT 18-SEP-1996 (first entry)  
DE Cw6 consensus peptide derived immunogenic peptide #2.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPT 96-116784/12.  
PT Compr. immunogenic peptide with supermotif allowing more  
than one HLA mol. to bind - used to induce CTL response in patient  
and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
use in the composition of the invention. The composition comprises  
an immunogenic peptide of 9-10 residues with a supermotif which  
allows binding of more than one HLA molecule. It pref. comprises  
two conserved residues, a first at the 2nd position from the N-  
terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
are used to induce a CTL response in a patient. They are also  
useful in compositions for in vivo and ex vivo therapeutic and  
diagnostic applications, e.g. the treatment of cancer and viral  
infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 100.0%; Score 70; DB 18; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.47e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 1 IYPYIVRS L 9  
QY 1 IYPYIVRS L 9

RESULT 2  
 ID R89369 standard; peptide; 9 AA.  
 AC R89369; (first entry)  
 DT 18-SEP-1996 (first entry)  
 DE Cw6 consensus peptide derived immunogenic peptide #1.  
 KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
 therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
 hepatitis C.  
 OS Synthetic.  
 PN W09603140-A1.  
 PD 08-FEB-1996.  
 PF 21-JUL-1995; U09234.  
 PR 21-JUL-1994; US-278634.  
 PR 23-NOV-1994; US-344824.  
 PR 30-MAY-1995; US-452843.  
 PA (CYTE-) CYTEL CORP.  
 PI Sette A, Sidney J;  
 WPI: 96-116784/12.  
 PT Compn. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications  
 PS Claim 2; Page 26; 32pp; English.  
 CC The sequences given in R89362-82 are immunogenic peptides which were used in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which allows binding of more than one HLA molecule. It pref. comprises two conserved residues, a first at the 2nd position from the N-terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also useful in compositions for in vivo and ex vivo therapeutic and diagnostic applications, e.g. hepatitis B and C.  
 CC infections, e.g. hepatitis B and C.  
 SO Sequence 9 AA; .

Query Match 92.9%; Score 65; DB 18; Length 9;  
 Best Local Similarity 88.9%; Pred. No. 6.41e-01;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 lpylrvl 9  
 |||||  
 1 lpylrvsl 9

RESULT 3  
 ID R89371 standard; peptide; 9 AA.  
 AC R89371;  
 DT 18-SEP-1996 (first entry)  
 DE Cw6 consensus peptide derived immunogenic peptide #3.  
 KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
 therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
 hepatitis C.  
 OS Synthetic.  
 PN W09603140-A1.  
 PD 08-FEB-1996.  
 PF 21-JUL-1995; U09234.  
 PR 21-JUL-1994; US-278634.  
 PR 23-NOV-1994; US-344824.  
 PR 30-MAY-1995; US-452843.  
 PA (CYTE-) CYTEL CORP.  
 PI Sette A, Sidney J;  
 WPI: 96-116784/12.  
 PT Compn. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications  
 PS Claim 2; Page 26; 32pp; English.  
 CC The sequences given in R89362-82 are immunogenic peptides which were used in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which allows binding of more than one HLA molecule. It pref. comprises two conserved residues, a first at the 2nd position from the N-terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also

CC useful in compositions for in vivo and ex vivo therapeutic and  
 CC diagnostic applications, e.g. the treatment of cancer and viral  
 CC infections, e.g. hepatitis B and C.  
 SO Sequence 9 AA;

Query Match 81.4%; Score 57; DB 18; Length 9;  
 Best Local Similarity 77.8%; Pred. No. 6.33e+00;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 1 lpylrvl 9  
 |||||  
 1 lpylrvsl 9

RESULT 4  
 ID W28511 standard; protein; 133 AA.  
 AC W28511;  
 DT 29-DEC-1997 (first entry)  
 DE Product of clone 1405.  
 KW J5; J422; I405; H174-10; H174-43; B18; cytokine; PMBC;  
 KW peripheral blood mononuclear cell; disintegrin; metallo-protein;  
 KW thiosphilia; leucine-rich repeat; monocyte; chemottractant;  
 KW Ie-10; CRG-2; CTLA-8; herpesvirus; Salmifl.  
 OS Mus musculus.  
 PN W09707198-A2.  
 PD 27-FEB-1997.  
 PF 08-AUG-1996; U12897.  
 PR 08-AUG-1996; WO-U12897.  
 PA (GENY ) GENETICS INST INC.  
 PI Carlin M, Jacobs K, Kelleher K, McCoy JM;  
 WPI: 97-165283/15.  
 DR N-PSDB: T87429.  
 PT Polynucleotide(s) encoding proteins for treating, preventing and ameliorating medical conditions - obtained from human activated peripheral blood mononuclear cell, and murine adult thymus libraries  
 PS Claim 21; Page 44-45; 61pp; English.  
 CC This sequence was isolated from a murine adult thymus library using a trip selecting for nucleotides encoding secreted proteins, and encodes a protein having homology to various monocyte and other chemottractant proteins.  
 SO Sequence 133 AA;

Query Match 74.3%; Score 52; DB 25; Length 133;  
 Best Local Similarity 75.0%; Pred. No. 2.53e+01;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 40 lpylrvl 47  
 |||||  
 1 lpylrvsl 8

RESULT 5  
 ID P70388 standard; protein; 356 AA.  
 AC P70388;  
 DT 14-JAN-1991 (first entry)  
 DE D-amino acid oxidase.  
 KW D-amino acid oxidase; Trigonopsis variabilis; cephalosporin;  
 KW oxidative deamination.  
 OS Trigonopsis variabilis.  
 PN J62262994-A.  
 PD 16-NOV-1987.  
 PF 12-MAY-1986; J06663.  
 PR 12-MAY-1986; JP-106663.  
 PA (ASAH ) ASAH CHEMICAL IND KK.  
 WPI: 87-359677/51.  
 DR N-PSDB: N70609.  
 PT DNA fragment encoding D-amino acid oxidase - which is a useful enzyme for the catalytic oxidative deamination of D-amino acids.  
 PS Claim 1; page 583-4; 12pp; Japanese.  
 CC D-amino acid oxidase catalyses the oxidative deamination of D-amino acids. It is used in the sepn. of L-amino acids from racemates, in the prepn. of ketoic acid from D-amino acid, in amino acid analysis, etc. The enzyme can oxidise cephalosporin C to

CC 7-beta-(5-carboxy-5-oxopentanamide)cephalosporanic acid, which  
 CC reacts with hydrogen peroxide to give 7-beta-(4-carboxypentanamide)-  
 CC cephalosporanic acid. These cpds. are important intermediates for  
 CC synthesis of cephalosporin type antibiotics.  
 SQ Sequence 356 AA;

Query Match 71.4%; Score 50; DB 2; Length 356;  
 Best Local Similarity 55.6%; Pred. No. 4.36e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 66 vsypilrel 74  
 0 1 IPYPIVRSLS 9

RESULT 6  
 ID R04066 standard; Protein; 356 AA.  
 AC R04066;  
 DT 03-SEP-1990 (first entry)  
 DE T-variabilis D-amino acid oxidase gene product.  
 KW D-amino acid oxidase; cephalosporin; cephem; E.coli.  
 OS Trigonopsis variabilis.  
 PN EP-364275-A.  
 PD 18-APR-1990.  
 PF 12-OCT-1989; 310483.  
 PR 13-OCT-1988; JP-260332.  
 PA (Fujii) Fujisawa Pharm KK.  
 PI Iisogai T, Ono H, Kojo H;  
 DR WPI; 90-117771/16.  
 PT D-amino acid oxidase, prodn.  
 PT by culture of E.coli transformants contg. expression vectors  
 PT originated from Fusarium solani M-0718.  
 PS Disclosure; Fig 9; 38pp; English.  
 CC E.coli transformed to express DNO, which catalyses the enzymatic  
 CC conversion of cephalosporin C to 7-beta-(5-carboxy-5-  
 CC oxopentanamido)cephalosporanic acid (keto-7ACA). 7ACA is an  
 CC important starting point for the production of cepham  
 CC antibiotics.  
 SQ Sequence 356 AA;

Query Match 71.4%; Score 50; DB 3; Length 356;  
 Best Local Similarity 55.6%; Pred. No. 4.36e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 65 vsypilrel 73  
 0 1 IPYPIVRSLS 9

RESULT 7  
 ID R77604 standard; Protein; 1676 AA.  
 AC R77604;  
 DT 15-MAR-1996 (first entry)  
 DE Pro-C5 polypeptide.  
 KW Complement C5; haemolysis; kidney; glomerulonephritis;  
 KW monoclonal antibody; antiinflammatory; antibody engineering;  
 KW humanised antibody.  
 OS Homo sapiens.  
 FH Key  
 FH peptide  
 FT 1..18  
 FT /label= Sig-peptide  
 FT 19..673  
 FT /label= Beta-chain  
 FT 673..674  
 FT cleavage\_site  
 FT 677..678  
 FT peptide  
 FT 674..677  
 FT label= Cleavage\_peptide  
 FT 678..1676  
 FT protein  
 FT /label= Alpha-chain  
 FT /note= amino acids 872-892 (854-874 of  
 FT the mature protein) comprise the KSSKS  
 FT epitope"  
 FT 678..751  
 FT peptide

FT /label= C5a  
 FT 751..752  
 FT /label= Convertase\_cleavage\_site  
 FT modified\_site  
 FT 911  
 FT /label= N-glycosylation\_site  
 FT modified\_site  
 FT 1115  
 FT /label= N-glycosylation\_site  
 FT modified\_site  
 FT 1630  
 FT /label= N-glycosylation\_site  
 PN W09529697-A1.  
 PD 09-NOV-1995.  
 PF 01-MAY-1995; U05688.  
 PR 02-MAY-1994; US-236208.  
 PA (ALEX-) ALEXION PHARM INC.  
 PI Evans MJ, Matis L, Mueller EE, Nye SH, Rollins S;  
 PI Rother RP, Springhorn J P, Squinto SP, Thomas TC;  
 PI Wang Y, Wilkins JA;  
 DR WPI; 95-392923/50.  
 PT Treating glomerulonephritis with antibody against complement C5  
 PT component - to inhibit complement induced cell lysis  
 PS Example 13; Page 82-92; 181pp; English.  
 CC The cDNA sequence of the complement C5 gene transcript predicts a  
 CC secreted pro-C5 precursor of 1676 amino acids (R77604). C5 is a  
 CC beta-globulin heterodimer thought to play a role in the pathogenesis  
 CC of glomerulonephritis (GN). Cleavage of the C5 alpha-chain  
 CC by a convertase enzyme generates anaphylatoxic C5a. Monoclonal  
 CC and humanised recombinant antibodies that recognise the alpha-chain  
 CC KSSKC epitope (R77605) block C5a generation, thereby reducing  
 CC glomerular inflammation and kidney dysfunction associated with GN.  
 SQ Sequence 1676 AA;

Query Match 71.4%; Score 50; DB 15; Length 1676;  
 Best Local Similarity 62.5%; Pred. No. 4.36e+01;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 829 lpyvsvrg 836  
 0 1 IPYPIVRS 8

RESULT 8  
 ID W25084 standard; Protein; 311 AA.  
 AC W25084;  
 DT 30-DEC-1997 (first entry)  
 DE Haemophilus influenzae htrb polypeptide.  
 KW Vaccine; htrb gene; Gram-negative bacterium; non-toxic mutant;  
 KW pathogen; endotoxin; diagnosis; passive immunisation.  
 OS Haemophilus influenzae strain 2019.  
 PN W09719688-A1.  
 PD 05-JUN-1997.  
 PF 27-NOV-1996; U18984.  
 PR 01-DEC-1995; US-565943.  
 PA (AMCY ) AMERICAN CYANAMID CO.  
 PA (REGC ) UNIV CALIFORNIA.  
 PA (IOWA ) UNIV IOWA RES FOUND.  
 PI Apicella MA, Arumugham R, Gibson BW, Lee N, Sunshine MG;  
 DR WPI; 97-310355/28.  
 DR N-PSDB; T79708.  
 PT New Gram-negative bacterial pathogen vaccines - comprising a htrb  
 PT mutant or an endotoxin isolated from an htrb mutant optionally  
 PT conjugated to a carrier protein.  
 PS Example 1; Page 61-62; 79pp; English.  
 CC This polypeptide comprises the htrb gene product (see also T79708)  
 CC of Haemophilus influenzae strain 2019. A claimed vaccine  
 CC formulation contains as an active ingredient an htrb mutant of a  
 CC Gram-negative bacterial pathogen (GNBP), endotoxin isolated from an  
 CC htrb mutant (A) of a GNBP, endotoxin isolated from an  
 CC to a carrier protein, or (A) which has been genetically engineered  
 CC to express at least one heterologous vaccine antigen, where (A)  
 CC lacks one or more secondary acyl chains of lipid A contained in the  
 CC GNBP resulting in reduced toxicity when compared to lipid A of the  
 CC GNBP. Also claimed is a method for producing endotoxin-specific  
 CC antisera for diagnostic assays, or for passive immunisation,

CC comprising immunising an individual with a vaccine formulation  
 CC comprising an active ingredient as above, and collecting antibodies  
 CC produced from the immunised individual.  
 SQ Sequence 311 AA;

Query Match 67.1%; Score 47; DB 25; Length 311;  
 Best Local Similarity 71.4%; Pred. No. 9.72e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 34 Ipyplir 40  
 :|||||:  
 :|||||:  
 QY 1 IPYPIVRS 7

RESULT 9  
 ID R65017 standard; Protein: 1082 AA.  
 AC R65017;

DE 30-SEP-1995 (first entry)  
 DE PRB2 retinoblastoma tumour suppressor protein.  
 KM Retinoblastoma; tumour suppressor; PRB2; diagnostic.  
 OS Homo sapiens.

FN Key Location/Qualifiers  
 FT misc\_difference 507..512  
 FT note="primer A (R65015) recognition site"

PD WO9505470-A.  
 PN 23-FEB-1995.  
 PF 12-AUG-1994; U09293.  
 PR 12-AUG-1993; US-106493.

PA (UTEM) UNITV TEMPLE.  
 PI Giordano A;  
 DR WPI: 95-098768/13.

DR N-PSDB: 082748.  
 PT DNA encoding retinoblastoma suppressor protein, PRB2 - and  
 PT recombinant cell lines, for the diagnosis and suppression of  
 PT cells infected with adenovirus E1A.

PS Claim 10; Page 19-22; 29pp; English.  
 CC The retinoblastoma tumour suppressor protein, PRB2, binds to the  
 CC E1A transforming domain, and is useful in the diagnosis of cells  
 CC infected with adeno virus E1A or a related virus. The protein  
 CC may be administered as a cell growth suppressor to infected cells  
 CC e.g. retinoblastoma interocular cancer cells, and may be useful  
 CC for identifying other DNA tumour virus oncoproteins.

SQ Sequence 1082 AA;  
 Query Match 67.1%; Score 47; DB 13; Length 1082;  
 Best Local Similarity 71.4%; Pred. No. 9.72e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 979 PYPFVRT 985  
 :|||||:  
 :|||||:  
 QY 2 PYPIVRS 8

RESULT 10  
 ID W17785 standard; Protein: 1230 AA.  
 AC W17785;

DE 16-AUG-1997 (first entry)  
 DE Potato tuber soluble starch synthase.  
 KM Starch synthase; transgenic plant; potato; rice; Oryza sativa;  
 KM tomato; Lycopersicon esculentum; wheat; Triticum aestivum; cassava;  
 KM Manihot esculenta; sweet potato; Ipomoea batatas; barley;

KM Hordeum vulgare; oat; Avena; maize; Zea mays.  
 OS Solanum tuberosum cv. Desiree.  
 FN Key Location/Qualifiers  
 FT peptide 1..60  
 FT label="Sig\_peptide"

EP-779363-A2.  
 PD 18-JUN-1997.  
 PF 11-DEC-1996; 309004.  
 PR 12-DEC-1995; GB-023533.

PA (NAT) NAT STARCH & CHEM INVESTMENT HOLDING COR.  
 PI Edwards EA, Marshall J, Martin CR, Smith AM;  
 DR WPI: 97-312737/29.

DR N-PSDB: T68646.  
 PT Soluble starch synthase - used to produce altered starch from  
 PT commercially important plants, e.g. potato, rice, wheat, and maize  
 PS Claim 15; Page 18-24; 39pp; English.  
 CC The amino acid sequence (W17785) of a novel soluble starch synthase  
 CC (SSS) was deduced from a cDNA clone (T68646) obtd. from potato  
 CC tubers. The SSS has a specific activity greater than that of other  
 CC starch synthases, and may represent a novel class of starch  
 CC synthases. DNA constructs containing SSS nucleic acids in sense  
 CC or antisense orientation relative to a plant promoter can be used  
 CC to suppress SSS activity in transgenic plants, esp. potato, tomato,  
 CC rice, wheat, pea, cassava, sweet potato, barley, oat or maize.  
 CC This has little effect on starch or amylose content, but alters  
 CC starch characteristics. A novel altered starch, extracted from  
 CC transgenic potato, has a viscosity onset temperature that is at  
 CC least 5 degC lower than that of starch from non-transformed plants.  
 SQ Sequence 1230 AA;

Query Match 67.1%; Score 47; DB 23; Length 1230;  
 Best Local Similarity 55.6%; Pred. No. 9.72e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 3 YPIPIHRS 11  
 :|||||:  
 :|||||:  
 QY 1 IYPIVRS 9

RESULT 11  
 ID W15422 standard; Protein: 225 AA.  
 AC W15422;

DE 05-JUN-1997 (first entry)  
 DE G protein conjugative receptor.  
 KM G protein conjugative receptor; GPCR; rabbit;  
 KM gastric pylorus smooth muscle.  
 OS Oryctolagus cuniculus.

PN J09051795-A.  
 PD 25-FEB-1997.  
 PF 11-AUG-1995; 205795.

PR 11-AUG-1995; JP-205795.  
 PA (TAKE) TAKEDA CHEM IND LTD.  
 DR WPI: 97-196269/18.

DR N-PSDB: T65490.  
 PT A new G protein conjugative receptor protein - isolated from rabbit  
 PT smooth muscle

PS Claim 1; Page 22; 25pp; Japanese.  
 CC This sequence represents G protein conjugative receptor (GPCR) protein.  
 CC The protein and the DNA can be used for genetic treatment. The cDNA  
 CC encoding this sequence was isolated from a polyA+ RNA fraction prepared  
 CC from rabbit gastric pylorus smooth muscle.

SQ Sequence 225 AA;  
 Query Match 65.7%; Score 46; DB 21; Length 225;  
 Best Local Similarity 55.6%; Pred. No. 1.26e+02;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 12 mpyamvrs 20  
 :|||||:  
 :|||||:  
 QY 1 IYPIVRS 9

RESULT 12  
 ID W35832 standard; Protein: 225 AA.  
 AC W35832;

DE 27-FEB-1998 (first entry)  
 DE Plasmid pUD-BL5 human G protein conjugate type receptor protein fragment.  
 KM Human; G protein; guanine nucleotide binding protein; gene therapy;  
 KM receptor ligand; genetic diagnosis.  
 OS Synthetic.

OS Homo sapiens.  
 PN J09238686-A.  
 PD 16-SEP-1997.  
 PF 07-MAR-1996; 050678.  
 PR 07-MAR-1996; JP-050678.



PA (TAKE ) TAKEDA CHEM IND LTD.  
 DR WPI: 97-506555/47.  
 DR N-PSDB: T96973.  
 PT Novel G protein conjugate receptor - used for identifying receptor  
 PS ligands which may potentially be useful in therapeutic drugs  
 PS Disclosure: Page 25-26; 31pp; Japanese.  
 CC The present sequence represents a G protein (guanine nucleotide-binding  
 CC protein) conjugate-type receptor protein fragment from plasmid pUD-BL5.  
 CC The G protein and its encoding DNA are used in the development of a  
 CC receptor-binding assay system for screening for candidate drugs. The  
 CC DNA and fragments of it may also be used as primers or probes for  
 CC genetic diagnosis, and in gene therapy. The elucidation of the structure  
 CC and properties of the G protein conjugate-type receptor is expected to  
 CC lead to the development of unique drugs acting on its system.  
 SO Sequence 225 AA;

Query Match 65.7%; Score 46; DB 26; Length 225;  
 Best Local Similarity 55.6%; Pred. No. 1.26e+02;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 12 mpysmrvs 20  
 :||:||||:  
 QY 1 IPYIVRSL 9

RESULT 13  
 ID W35346 standard; Protein: 524 AA.  
 AC W35346;  
 DT 14-APR-1998 (first entry)  
 DE Arabidopsis thaliana epsilon cyclase.  
 KW Epsilon cyclase; E-cyclase; carotenoid; biosynthetic enzyme;  
 KW pigment; vector; pATeps.  
 OS Arabidopsis thaliana.  
 PN W09736998-A1.  
 PD 09-OCT-1997.  
 PF 28-JAN-1997; U00540.  
 PR 29-MAR-1996; US-624125.  
 PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 PI Cunningham FX, Sun Z;  
 DR WPI: 97-503091/46.  
 DR N-PSDB: T95371.  
 PT Eukaryotic carotenoid biosynthetic enzymes and related genes -  
 PT useful to control ratio of various carotenoid(s) in host and for  
 PT production of novel carotenoid pigments  
 PS Claim 1, Page 33-35; 89pp; English.  
 CC This protein comprises Arabidopsis thaliana epsilon cyclase, an  
 CC enzyme responsible for the formation of epsilon end-groups in  
 CC carotenoids. Its amino acid sequence was deduced from an  
 CC isolated cDNA clone (see T95371). Expression vector pATeps  
 CC comprising the epsilon cyclase gene is deposited as ATCC 98005.  
 CC The claimed eukaryotic carotenoid biosynthetic enzymes  
 CC epsilon cyclase, beta-carotene hydroxylase and isopentenyl  
 CC pyrophosphate isomerase (see W35346-51) are used in methods for  
 CC augmenting the accumulation of carotenoids and for the production  
 CC of novel and rare carotenoids in host cells. Methods are also  
 CC provided for controlling the ratio of various carotenoids in a  
 CC host, and for screening for eukaryotic genes that encode enzymes  
 CC of carotenoid biosynthesis and metabolism.  
 SO Sequence 524 AA;

Query Match 65.7%; Score 46; DB 27; Length 524;  
 Best Local Similarity 71.4%; Pred. No. 1.26e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 390 ysvvrs1 396  
 :||:||||:  
 QY 3 YPIVRS1 9

RESULT 14  
 ID R15444 standard; Protein: 1330 AA.  
 AC R15444;  
 DT 16-MAR-1992 (first entry)

DE Swine herpes virus-1 major capsid protein.  
 KW MCP; p1g; recombinant vaccinia virus; SHV-1; PHAS2-MCP; mad itch.  
 OS Pseudorabies virus.  
 PN J03247285-A.  
 PD 05-NOV-1991.  
 PF 27-FEB-1990; 046888.  
 PR 27-FEB-1990; JP-046888.  
 PA (NORO) NORINSHO.  
 PA (MITU) MITSUBISHI KASEI CORP.  
 DR WPI: 91-366332/50.  
 DR N-PSDB: Q15153.  
 PT Capsid protein of swine herpes virus type I - used for the  
 PT diagnosis, prevention and treatment of Aujeszky disease  
 PS Claim 4, Fig 1, 22pp; Japanese.  
 CC The SHV-1 MCP can be efficiently produced by recombinant DNA  
 CC technology. See also Q15154.  
 SO Sequence 1330 AA;

Query Match 65.7%; Score 46; DB 3; Length 1330;  
 Best Local Similarity 55.6%; Pred. No. 1.26e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 368 vpyplvgn1 376  
 :||:||||:  
 QY 1 IPYIVRSL 9

RESULT 15  
 ID R13499 standard; Protein: 210 AA.  
 AC R13499;  
 DT 25-OCT-1991 (first entry)  
 DE P.dentriticans COB H.  
 KW cob gene; corrinoid; desiccatochorinoid; cor gene.  
 OS Pseudomonas dentriticans.  
 PN W0911518-A.  
 PD 08-AUG-1991.  
 PF 30-JAN-1991; F00054.  
 PR 31-JAN-1990; FR-001137.  
 PA (RHON) RHONE-POULENC BIOCH.  
 PI Blanche F, Meron B, Crouzet J, Debussche L, Levy-Schil S;  
 PI Thibaut D.  
 DR WPI: 91-252650/34.  
 DR N-PSDB: Q13285.  
 PT New polypeptide(s) involved in cobalamin and cobamide  
 PT biosynthesis - and DNA encoding them, for amplification of  
 PT cobalamin, esp. coenzyme B12 prodn.  
 PS Claim 25; Fig 16; 299pp; French.  
 CC This sequence corresponds to one of 24 polypeptides obtained from  
 CC P.dentriticans and implicated in the biosynthesis of cobalamines  
 CC and/or cobamides. It has precorin-8x mutase activity. It is encoded  
 CC by part of the 8.7kb EcoRI-EcoRI fragment of plasmid pX1367. The  
 CC plasmid was isolated from a P.dentriticans genomic DNA bank  
 CC constructed in vector pX159.  
 CC See Q13284-Q13288.  
 SO Sequence 210 AA;

Query Match 64.3%; Score 45; DB 3; Length 210;  
 Best Local Similarity 50.0%; Pred. No. 1.64e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 183 vpfatvg 190  
 :||:||||:  
 QY 1 IPYIVRS 8

Search completed: Fri Sep 11 13:09:49 1998  
 Job time : 14 secs.

**THIS PAGE BLANK (USPTO)**

# MIRAGE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mprch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Sep 11 13:10:08 1998; Maspar time 3.17 Seconds  
Tabular output not generated. 103.708 Million cell updates/sec

Title: >US-08-452-843-9  
Description: (1-9) from US08452843.pep  
Perfect Score: 70  
Sequence: 1 IPRYVRS L 9

Scoring table: PAM 150  
Gap 15

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: p1r56  
1:p1r1 2:p1r2 3:p1r3 4:p1r4 5:nr13d

Statistics: Mean 23.565; Variance 32.285; scale 0.731

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description            | Pred. No. |
|------------|-------|-------------|--------|----|--------|------------------------|-----------|
| 1          | 56    | 80.0        | 340    | 2  | S62493 | hypothetical protein   | 1.21e+00  |
| 2          | 53    | 75.7        | 209    | 2  | S13179 | transforming protein   | 4.57e+00  |
| 3          | 53    | 75.7        | 487    | 1  | COEGRS | hypothetical Y9J6 pro  | 1.08e+01  |
| 4          | 51    | 72.9        | 381    | 2  | C64416 | hypothetical protein   | 1.08e+01  |
| 5          | 51    | 72.9        | 478    | 2  | F69811 | 2-oxoglutarate/malate  | 1.08e+01  |
| 6          | 50    | 71.4        | 1061   | 1  | S27311 | ribonuclease E (EC 3.  | 1.65e+01  |
| 7          | 50    | 71.4        | 1676   | 1  | CSHU   | complement C5 precurs  | 1.65e+01  |
| 8          | 50    | 71.4        | 1680   | 1  | CSMS   | complement C5 precurs  | 1.65e+01  |
| 9          | 50    | 71.4        | 4344   | 2  | A53489 | cytoplasmic dynein he  | 1.65e+01  |
| 10         | 50    | 71.4        | 4367   | 2  | B54802 | dynein heavy chain, c  | 2.50e+01  |
| 11         | 49    | 70.0        | 57     | 2  | D35826 | hypothetical protein   | 2.50e+01  |
| 12         | 49    | 70.0        | 219    | 2  | S75541 | hypothetical protein   | 2.50e+01  |
| 13         | 49    | 70.0        | 238    | 2  | B64313 | hypothetical protein   | 2.50e+01  |
| 14         | 49    | 70.0        | 298    | 1  | MMAGCF | membrane protein lacF  | 2.50e+01  |
| 15         | 48    | 68.6        | 78     | 2  | C64472 | hypothetical protein   | 3.78e+01  |
| 16         | 48    | 68.6        | 303    | 2  | JN0857 | C alpha dehydrogenase  | 3.78e+01  |
| 17         | 48    | 68.6        | 305    | 2  | S35991 | conserved hypothetical | 3.78e+01  |
| 18         | 48    | 68.6        | 432    | 1  | G69993 | transcription factor   | 3.78e+01  |
| 19         | 48    | 68.6        | 440    | 1  | TWIM1  | transcription factor   | 3.78e+01  |
| 20         | 48    | 68.6        | 451    | 2  | S75569 | DNA-directed RNA poly  | 3.78e+01  |
| 21         | 48    | 68.6        | 880    | 2  | B33926 | hypothetical protein   | 5.66e+01  |
| 22         | 48    | 68.6        | 1026   | 2  | S51432 | hypothetical protein   | 5.66e+01  |
| 23         | 47    | 67.1        | 228    | 2  | S75394 | hypothetical protein   | 5.66e+01  |

|    |    |      |      |   |         |                        |          |
|----|----|------|------|---|---------|------------------------|----------|
| 24 | 47 | 67.1 | 230  | 2 | S71747  | DAG protein precursor  | 5.66e+01 |
| 25 | 47 | 67.1 | 232  | 2 | S28795  | carbonate dehydratase  | 5.66e+01 |
| 26 | 47 | 67.1 | 302  | 2 | S71730  | hypothetical protein   | 5.66e+01 |
| 27 | 47 | 67.1 | 302  | 2 | S60955  | probable membrane pro  | 5.66e+01 |
| 28 | 47 | 67.1 | 315  | 2 | D64127  | htrb protein - Haemop  | 5.66e+01 |
| 29 | 47 | 67.1 | 356  | 1 | JQ1582  | major capsid protein   | 5.66e+01 |
| 30 | 47 | 67.1 | 364  | 2 | S77360  | hypothetical protein   | 5.66e+01 |
| 31 | 47 | 67.1 | 434  | 2 | G70011  | conserved hypothetical | 5.66e+01 |
| 32 | 47 | 67.1 | 634  | 2 | S33575  | dnak-type molecular c  | 5.66e+01 |
| 33 | 47 | 67.1 | 675  | 2 | S19140  | probable membrane pro  | 5.66e+01 |
| 34 | 47 | 67.1 | 1056 | 2 | S55151  | Rb2/P130 protein - hu  | 5.66e+01 |
| 35 | 47 | 67.1 | 1082 | 2 | I38150  | adenovirus E1A-associ  | 5.66e+01 |
| 36 | 47 | 67.1 | 1139 | 2 | A49369  | E1A-associated cyclin  | 5.66e+01 |
| 37 | 47 | 67.1 | 1139 | 2 | A49370  | hypothetical protein   | 8.43e+01 |
| 38 | 47 | 67.1 | 1679 | 2 | S48385  | ribosomal protein l21  | 8.43e+01 |
| 39 | 46 | 65.7 | 118  | 2 | H64304  | agmatinase (EC 3.5.3.  | 8.43e+01 |
| 40 | 46 | 65.7 | 284  | 2 | F64338  | conserved hypothetical | 8.43e+01 |
| 41 | 46 | 65.7 | 336  | 2 | H64610  | hypothetical protein   | 8.43e+01 |
| 42 | 46 | 65.7 | 433  | 2 | S77395  | glucose-6-phosphate 1  | 8.43e+01 |
| 43 | 46 | 65.7 | 607  | 1 | NU07B   | helicase (EC 3.6.1.-)  | 8.43e+01 |
| 44 | 46 | 65.7 | 1174 | 1 | HJBYDH  | major capsid protein   | 8.43e+01 |
| 45 | 46 | 65.7 | 1376 | 1 | VCBERD6 |                        |          |

## ALIGNMENTS

| RESULT ENTRY TITLE    | 1  | 2              |
|-----------------------|--|----------------|
| S62493                | hypothetical protein SPAC23D3.02 - fission yeast (Schizosaccharomyces pombe)                         |                |
| ORGANISM              | #formal_name Schizosaccharomyces pombe   |                |
| DATE                  | 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 31-Oct-1997                                  |                |
| ACCESSIONS            | S62493   |                |
| REFERENCE             | Niblett, D.; Harris, D. submitted to the EMBL Data Library, October 1995                             |                |
| #submissions          | S62493   |                |
| #accession            | S62493   |                |
| #status               | preliminary  |                |
| #molecule_type        | DNA  |                |
| #residues             | 1-340 #label NIB   |                |
| #cross-references     | EMBL:Z64354; NID:g1039338; PID:g1039340  |                |
| GENETICS              | #map_position 1R   |                |
| #introns              | 25/2   |                |
| SUMMARY               | #length 340 #molecular-weight 37876 #checksum 2036   |                |
| Query Match           | 80.0%; Score 56; DB 2; Length 340;   |                |
| Best Local Similarity | 66.7%; Pred. No. 1.21e+00;   |                |
| Matches               | 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;   |                |
| DB                    | 249 VPRYIRSL 257   |                |
| QY                    | 1 IPRYVRS L 9  |                |
| RESULT                | 2  |                |
| ENTRY                 | S13179   | #type complete |
| TITLE                 | transforming protein (ras) - Geodia cydonium   |                |
| ORGANISM              | #formal_name Geodia cydonium   |                |
| DATE                  | 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997                                  |                |
| ACCESSIONS            | S13179   |                |
| REFERENCE             | S13179   |                |
| #authors              | Robitzki, A.; Schroeder, H.C.; Ugarkovic, D.; Kuchino, Y.; Kurelec, B.; Gamulin, V.; Mueller, W.E.G. |                |
| #journal              | Eur. J. Biochem. (1990) 192:499-506  |                |
| #title                | Regulated expression and phosphorylation of the 23-26-kDa ras protein in the sponge Geodia cydonium. |                |
| #cross-references     | NUID:g1006138  |                |
| #accession            | S13179   |                |
| #status               | preliminary  |                |
| #molecule_type        | mRNA   |                |

```

#residues      1-209 ##label ROB
#note          based on the evidence for Gln-tRNA, the authors
               translated the codon TAG as Gln; the sequence shown
               follows the authors' translation
CLASSIFICATION #superfamily ras transforming protein
SUMMARY        #length 209 #molecular-weight 23854 #checksum 3860

Query Match    75.7%; Score 53; DB 2; Length 209;
Best Local Similarity 66.7%; Pred. No. 4.57e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db: 1 178 IPYSLVREL 186
1 IPYSLVREL 9

RESULT 3
ENTRY      OOECRS      #type complete
TITLE      hypothetical yjgE protein - Escherichia coli (strain K-12)
ORGANISM   Escherichia coli
DATE       30-Jun-1988 #sequence-revision 31-Oct-1997 #text-change
ACCESSIONS E65094; C29049
REFERENCE   A64720
            Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
            Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
            Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
            Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
            Y.
            Science (1997) 277:1453-1462
            #title The complete genome sequence of Escherichia coli K-12.
            #cross-references GB:AE000388; GB:U00096; NID:g1789441; PID:g1789444;
            #accession E65094
            ##status nucleic acid sequence not shown; translation not shown
            ##molecule-type DNA
            ##residues 1-487 ##label BLAT
            ##cross-references GB:AE000388; GB:U00096; NID:g1789441; PID:g1789444;
            #accession UMCB:B3063
            ##experimental source strain K-12, substrain MG1655
REFERENCE   A91573
            Nesin, M.; Lupski, J.R.; Svec, P.; Godson, G.N.
            Gene (1987) 51:149-161
            #journal Possible new genes as revealed by molecular analysis of a
            #title 5-kb Escherichia coli chromosomal region 5' to the
            ipu-dna-gripod macromolecular-synthesis operon.
            #cross-references NID:87248073
            #accession C29049
            ##molecule-type DNA
            ##residues 279-403, 'P', #405-411, 'RWRCRKRRCRCSA' ##label NES
GENETICS
#gene        yjgE
#map-position 67 min
#length 487 #molecular-weight 52906 #checksum 1643
SUMMARY
Query Match    75.7%; Score 53; DB 1; Length 487;
Best Local Similarity 66.7%; Pred. No. 4.57e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db: 165 IIPYLRNL 173
1 IIPYLRNL 9

RESULT 4
ENTRY      C64416      #type complete
TITLE      hypothetical protein MG372 homolog - Methanococcus jannaschii
ORGANISM   Methanococcus jannaschii
DATE       13-Sep-1996 #sequence-revision 13-Sep-1996 #text-change
ACCESSIONS C64416
            A64300
            Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
            R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,

```

```

R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrman,
J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
Science (1996) 273:1058-1073
#journal      Complete genome sequence of the methanogenic archaeon,
#title        Methanococcus jannaschii.
#cross-references NID:96337999
#accession    C64416
#status       preliminary; nucleic acid sequence not shown;
               translation not shown
            ##molecule-type DNA
            ##residues 1-381 ##label BUL
            ##cross-references GB:U67536; GB:L77117; NID:g1591596; PID:g1591602;
            TIGR:MJ0931; PID:g1510973
SUMMARY      #map-position FOR860923-862068
               #length 381 #molecular-weight 43436 #checksum 7754
GENETICS
#gene        yfjS
#map-position 72.9%; Score 51; DB 2; Length 381;
Best Local Similarity 66.7%; Pred. No. 1.08e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db: 319 INPYLRPL 327
1 INPYLRPL 9

RESULT 5
ENTRY      F69811      #type complete
TITLE      2-oxoglutarate/malate translocator homolog yfjS - Bacillus
            subtilis
ORGANISM   Bacillus subtilis
DATE       05-Dec-1997 #sequence-revision 05-Dec-1997 #text-change
ACCESSIONS F69811
            A69580
            Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
            Alloni, G.; Azevedo, V.; Berto, M.G.; Bessieres, P.;
            Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
            A.; Braun, M.; Brgnelli, S.C.; Bron, S.; Brouillet, S.;
            Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
            Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.;
            Daniel, R.A.; Denzot, F.; Devine, K.M.; Duesterhoeft, A.;
            Ehrlich, S.D.; Emerson, P.T.; Enliar, K.D.; Errington, J.;
            Fabret, C.; Ferrari, E.; Folger, D.; Fritz, C.; Fujita,
            M.; Fujita, Y.; Fuma, S.; Galliz, A.; Galleron, N.; Gilm,
            S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;
            Giuseppe, G.; Guy, B.J.; Haga, A.; Gollightly, E.J.; Harwood,
            C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
            Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
            Kasahara, Y.; Klier-Bianchard, M.; Klein, C.; Kobayashi,
            Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
            Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
            Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
            Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
            M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
            V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
            A.M.; Presecan, E.; Pulic, P.; Punelle, B.; Rapoport, G.;
            Rey, M.; Reynolds, S.; Rieger, P.; Rivolta, C.; Rocha, E.;
            Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
            Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
            Sekowska, A.; Seror, S.J.; Serori, P.; Shin, B.S.; Soldo,
            B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
            Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.;
            Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
            Vandenbol, M.; Vannier, F.; Vassartotti, A.; Viari, A.;
            Wambutt, R.; Wedler, E.; Wedler, H.; Weltzenegger, T.;

```

Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
 #journal Nature (1997) 390:249-256  
 #title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 #accession F69811  
 #status preliminary; nucleic acid sequence not shown; translation not shown  
 ##molecule\_type DNA  
 ##residues 1-478 ##label KUN  
 ##experimental\_source strain 168  
 GENETICS  
 #gene yfjS  
 SUMMARY #length 478 #molecular-weight 51431 #checksum 768  
 Query Match 72.9%; Score 51; DB 2; Length 478;  
 Best Local Similarity 66.7%; Pred. No. 1.08e+01;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 DB 155 IFFPIRSL 163  
 1-1111111111  
 1 IPIPIVRS 9  
 Oy  
 RESULT 6  
 ENTRY S27311 #type complete  
 TITLE ribonuclease E (EC 3.1.4.-) - *Escherichia coli*  
 ALTERNATE\_NAMES cell shape determining protein; message stability-altering protein; RNase E  
 ORGANISM #formal\_name *Escherichia coli*  
 DATE 30-Sep-1993 #sequence\_revision 05-Dec-1997 #text\_change 05-Dec-1997  
 A64852; S45572; S27311; A23747; JG0009; A40661; S13127;  
 S25116  
 REFERENCE A64720  
 #authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 #journal Science (1997) 277:1453-1462  
 #title The complete genome sequence of *Escherichia coli* K-12.  
 #cross-references MIMD:97426617  
 #accession A64852  
 #status nucleic acid sequence not shown; translation not shown  
 ##molecule\_type DNA  
 ##residues 1-1061 ##label BLAT  
 ##cross-references GB:AE000209; GB:U00096; NID:q1787322; PID:q1787325; UMGF:bl084  
 #experimental\_source strain K-12, substrain MG1655  
 S45572  
 REFERENCE  
 #authors Casaregola, S.; Jacq, A.; Laoudj, D.; McGurk, G.; Margaron, S.; Tempete, M.; Norris, V.; Holland, I.B.  
 #journal J. Mol. Biol. (1994) 238:867  
 #title Cloning and analysis of the entire *Escherichia coli* *ams* gene.  
 #accession S45572  
 ##molecule\_type DNA  
 ##residues 1001-1061 ##label CAS  
 S27311  
 REFERENCE  
 #authors Casaregola, S.; Jacq, A.; Laoudj, D.; McGurk, G.; Margaron, S.; Tempete, M.; Norris, V.; Holland, I.B.  
 #journal J. Mol. Biol. (1992) 228:30-40  
 #title Cloning and analysis of the entire *Escherichia coli* *ams* gene. *ams* is identical to *hmp1* and encodes a 114 kDa protein that migrates as a 180 kDa protein.  
 #accession S27311  
 ##molecule\_type DNA  
 ##residues 1-486, 'V', 488-563, 'R', 565-783, 'K', 785-904, 'R', 906-1000, 1060-1061, 'ITLPPADASSGICSGANASQ' ##label CA2  
 A23747  
 REFERENCE  
 #cross-references EMBL:X67470; NID:q49115; PID:q49116  
 A23747  
 #authors Claverie-Martin, F.; Diaz-Torres, M.R.; Yancey, S.D.;

Kushner, S.R.  
 #journal J. Biol. Chem. (1991) 266:2843-2851  
 #title Analysis of the altered mRNA stability (*ams*) gene from *Escherichia coli*. Nucleotide sequence, transcriptional analysis, and homology of its product to MRP3, a mitochondrial ribosomal protein from *Neurospora crassa*.  
 #cross-references MIMD:91131576  
 #accession A23747  
 #status preliminary  
 ##molecule\_type DNA  
 ##residues 1-389, 'H', 391-486, 'V', 488-795, 'SF', 798, 1009, 'LASS', 1014-1015, 'RKMSASSLS' ##label CIA  
 ##cross-references GB:M62747; NID:q145271; PID:q145273  
 ##note this sequence has been proven to be erroneous in Ref:S27311  
 REFERENCE JG0009  
 #authors Chauhan, A.K.; Miczak, A.; Taraseviciene, L.; Agrillon, D.  
 #journal Nucleic Acids Res. (1991) 19:125-129  
 #title Sequencing and expression of the *rne* gene of *Escherichia coli*.  
 #cross-references MIMD:91187608  
 #accession JG0009  
 #status preliminary  
 ##molecule\_type DNA  
 ##residues 1-258, 'N', 260-529, 'QPLPCR', 'MC', 719, 'LR', 722-726, 'LPRLL', ##label CHA  
 ##cross-references EMBL:X54309  
 ##note this sequence has been proven to be erroneous in Ref:S27311  
 REFERENCE A40661  
 #authors McDowell, K.J.; Hernandez, R.G.; Lin-Chao, S.; Cohen, S.N.  
 #journal J. Bacteriol. (1993) 175:4245-4249  
 #title The *ams*-1 and *rne*-3071 temperature-sensitive mutations in the *ams* gene are in close proximity to each other and cause substitutions within a domain that resembles a product of the *Escherichia coli* *mare* locus.  
 #cross-references MIMD:93308106  
 #accession A40661  
 #status not compared with conceptual translation  
 ##molecule\_type DNA  
 ##residues 1-486, 'V', 488-489 ##label MCD  
 ##note sequence extracted from NCBI backbone (NCBIP:134520)  
 GENETICS  
 #gene *rne*  
 #map\_position 24 min  
 CLASSIFICATION #superfamily ribonuclease E  
 KEYWORDS endonuclease; hydrolase; P-loop; phosphoric diester hydrolase; RNA binding; transmembrane protein  
 FEATURE 113-131  
 113-176  
 169-176  
 524-568  
 743-778  
 SUMMARY #length 1061 #molecular-weight 118196 #checksum 5236  
 Query Match 71.4%; Score 50; DB 1; Length 1061;  
 Best Local Similarity 75.0%; Pred. No. 1.65e+01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 DB 844 IRIPIVRS 851  
 1-1111111111  
 1 IPIPIVRS 8  
 Oy  
 RESULT 7  
 ENTRY C5HT #type complete  
 TITLE complement C5 precursor - human  
 CONTAINS C5a anaphylatoxin; C5b  
 ORGANISM #formal\_name *Homo sapiens* #common\_name man  
 DATE 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 24-Oct-1997  
 A40075; A27689; A01267; A01266; S15121

REFERENCE A40075  
#authors Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Hunt, A.; Wetsel, R.A.  
#journal J. Immunol. (1991) 146:362-368  
#title Complete cDNA sequence of human complement pro-C5. Evidence of truncated transcripts derived from a single copy gene.  
#cross-references MIMD:91079575  
#accession A40075  
##molecule-type mRNA  
##residues 1-1676 ##label HAV  
#cross-references GB:M57729; NID:9179982; PID:9179983  
#note 518-Ser was also found

REFERENCE A27689  
#authors Wetsel, R.A.; Lemons, R.S.; Le Beau, M.M.; Barnum, S.R.; Noack, D.; Tack, B.F.  
#journal Biochemistry (1988) 27:1474-1482  
#title Molecular analysis of human complement component C5: localization of the structural gene to chromosome 9.  
#cross-references MIMD:88209511  
#accession A27689  
##molecule-type mRNA  
##residues 412-1676 ##label WET  
#cross-references GB:M6134; GB:M1879; NID:9179691; PID:9179692

REFERENCE A01267  
#authors Fernandez, H.N.; Hugli, T.E.  
#journal J. Biol. Chem. (1978) 253:6955-6964  
#title Primary structural analysis of the polypeptide portion of human C5a anaphylatoxin. Polypeptide sequence determination and assignment of the oligosaccharide attachment site in C5a.  
#cross-references MIMD:79005687  
#accession A01267  
##molecule-type protein  
##residues 678-751 ##label FER

REFERENCE A01266  
#authors Lundwall, A.B.; Wetsel, R.A.; Kristensen, T.; Whitehead, A.S.; Woods, D.E.; Ogden, R.C.; Colten, H.R.; Tack, B.F.  
#journal J. Biol. Chem. (1985) 260:2108-2112  
#title Isolation and sequence analysis of a cDNA clone encoding the fifth complement component.  
#cross-references MIMD:85130937  
#accession A01266  
##molecule-type mRNA  
##residues 412-854,  
'SIATSPRECKNGKISGCKRLRGSSDPASASQVAGITGHHNQPT',  
##label IDN  
#cross-references GB:K02874  
#note the carboxyl-terminal part of the sequence in this report appears to be derived from translation of an ALT repeat sequence

REFERENCE S15121  
#authors Bohnsack, J.F.; Mollison, K.W.; Buko, A.M.; Ashworth, J.C.; Hill, R.R.  
#journal Blochem. J. (1991) 273:635-640  
#title Group B streptococci inactivate complement component C5a by enzymic cleavage at the C-terminus.  
#cross-references MIMD:91144547  
#contents annotation

COMMENT Complement C5 contains two disulfide-linked chains, formed by removal of four basic residues. C5 convertase releases C5a anaphylatoxin from the amino end of the alpha chain, generating C5b (beta and alpha' chains).  
Activation of C5 initiates the spontaneous assembly of the late complement components, C5-C9, into the membrane attack complex. C5b has a transient binding site for C6. The C5b-C6 complex is the foundation upon which the membrane attack complex is assembled.

COMMENT C5a has potent spasmogenic and chemotactic activity.

GENETICS  
#gene GDB:C5  
#cross-references GDB:119734; OMIM:120900  
#map\_position 9q34.1-9q34.1  
CLASSIFICATION #superfamily alpha-2-macroglobulin

KEYWORDS complement alternate pathway; complement pathway; cytolysis; glycoprotein; inflammation; membrane attack complex; plasma

FEATURE  
1-18  
19-673, 678-1676  
19-673, 752-1676  
19-673  
678-1676  
678-751  
752-1676  
567-810, 634-669,  
698-724, 699-731,  
711-732, 866-1527,  
1101-1159,  
1375-1505,  
1405-1474,  
1520-1525,  
1532-1606,  
1533-1676,  
1654-1657  
741  
751-752  
911, 1115, 1630

SUMMARY #length 1676 #molecular-weight 188330 #checksum 3858

Query Match 71.4%; Score 50; DR 1; Length 1676;  
Best Local Similarity 62.5%; Pred. NO. 1,65e+01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 829 IPRSVYRG 836  
QY 1 IPRVIRS 8

RESULT 8  
ENTRY C5MS  
TITLE complement C5 precursor - mouse  
CONTAINS C5a anaphylatoxin; C5b  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 19-Nov-1988 #sequence\_revision 15-Oct-1994 #text\_change 24-Oct-1997

ACCESSIONS A35530; A27538; A40429  
REFERENCE A35530  
#authors Wetsel, R.A.; Fleischer, D.T.; Haviland, D.L.  
#journal J. Biol. Chem. (1990) 265:2435-2440  
#title Deficiency of the murine fifth complement component (C5). A 2-base pair gene deletion in a 5'-exon.  
#cross-references MIMD:90153853  
#accession A35530  
##molecule-type mRNA  
##residues 1-215, 'L' ##label WET  
#cross-references GB:M35526; GB:J05234; NID:9193302; PID:9309123  
REFERENCE A27538  
#authors Wetsel, R.A.; Ogata, R.T.; Tack, B.F.  
#journal Biochemistry (1987) 26:737-743  
#title Primary structure of the fifth component of murine complement.  
#cross-references MIMD:87185363  
#accession A27538  
##molecule-type mRNA  
##residues 'PGL' 44-1680 ##label WET2

REFERENCE A40429  
#authors Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Wetsel, R.A.; J. Biol. Chem. (1991) 266:11818-11825  
#journal Structure of the murine fifth complement component (C5) gene. A large, highly interrupted gene with a variant donor splice site and organizational homology with the third and

## Fourth complement component genes.

##cross-references MUID:91268053  
#accession A40429

##molecule\_type DNA  
##residues 1-15 #label HAV  
##cross-references GB:M64852

COMMENT Complement C5 contains two disulfide-linked chains, formed by removal of four basic residues. C5 convertase releases C3a anaphylatoxin from the amino end of the alpha chain, generating C5b (beta and alpha' chains).

COMMENT Activation of C5 initiates the spontaneous assembly of the late complement components, C5-C9, into the membrane attack complex. C5b has a transient binding site for C6. The C5b-C6 complex is the foundation upon which the membrane attack complex is assembled.

COMMENT C3a has potent spasmogenic and chemotactic activity.

## GENETICS

##map\_position 2  
#introns 22/3; 86/3; 140/3; 195/2; 223/1; 253/2; 291/3; 334/1;

## CLASSIFICATION

##superfamily alpha-2-macroglobulin  
complement alternate pathway; complement pathway: cytolysis; glycoprotein; inflammation; membrane attack complex; plasma

## FEATURE

1-18  
19-674, 679-1679  
19-674, 756-1679  
19-674  
679-1679

##domain signal sequence #status predicted #label SIG  
#product complement C5 #status predicted #label MAR  
#product C5b #status predicted #label C5b  
#product complement C5 and C5b beta chain #status  
#product complement C5 alpha chain #status predicted  
#label C5a  
#product C5a anaphylatoxin #status predicted #label C5a  
#product C5b alpha' chain #status predicted #label C5b

679-755  
756-1679  
567-814, 635-670,  
702-728, 703-735,  
713-736, 870-1531,  
1105-1163,  
1379-1509,  
1409-1478,  
1524-1529,  
1536-1609,  
1557-1679,  
1657-1660  
915,1119,1633

## SUMMARY

##disulfide\_bonds #status predicted  
#binding\_site carbohydrate (Asn) (covalent) #status  
predicted  
#length 1680 #molecular\_weight 18876 #checksum 3868

Query Match 71.4%; Score 50; DB 1; Length 1680;  
Best Local Similarity 62.5%; Pred. No. 1.65e+01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db: 833 IPYSVRG 840  
1 IPYIVRS 8

## RESULT

ENTRY 9  
TITLE cytoplasmic dynein heavy chain - Emertella nidulans  
ORFANISM 02-Jun-1994 #sequence\_revision 02-Jun-1994 #text\_change  
09-Sep-1997

ACCESSIONS A53489  
REFERENCE A53489  
#authors Xiang, X.; Beckwith, S.M.; Morris, N.R.  
#journal Proc. Natl. Acad. Sci. U.S.A. (1994) 91:2100-2104  
#title Cytoplasmic dynein is involved in nuclear migration in Aspergillus nidulans.  
#accession A53489

##status preliminary

##molecule\_type DNA  
##residues 1-4344 #label XIA  
##cross-references GB:U03904; NID:9451538; PID:9451539  
SUMMARY #length 4344 #molecular\_weight 492476 #checksum 8396

Query Match 71.4%; Score 50; DB 2; Length 4344;  
Best Local Similarity 75.0%; Pred. No. 1.65e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db: 394 PYPIRRAL 401  
2 PYPIVRS 9

## RESULT

ENTRY 10  
TITLE B54802 #type complete  
ORGANISM dynein heavy chain, cytoplasmic - Neurospora crassa  
DATE 23-Mar-1995 #sequence\_revision 05-Apr-1995 #text\_change  
09-Sep-1997

ACCESSIONS B54802  
REFERENCE A54802  
#authors Plamann, M.; Minke, P.F.; Tinsley, J.H.; Bruno, K.S.  
#journal J. Cell Biol. (1994) 127:139-149  
#title Cytoplasmic dynein and actin-related protein Arp1 are required for normal nuclear distribution in filamentous fungi.

#accession B54802  
##status preliminary

##molecule\_type DNA  
##residues 1-4367 #label PLA  
##cross-references GB:L31504; NID:9473489; PID:9473490

SUMMARY #length 4367 #molecular\_weight 495574 #checksum 8268

Query Match 71.4%; Score 50; DB 2; Length 4367;  
Best Local Similarity 75.0%; Pred. No. 1.65e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db: 404 PYPIRRAL 411  
2 PYPIVRS 9

## RESULT

ENTRY 11  
TITLE D35826 #type complete  
ORFANISM hypothetical protein B, 6.8K - human  
DATE 14-Dec-1990 #sequence\_revision 14-Dec-1990 #text\_change  
31-Oct-1997

ACCESSIONS D35826  
REFERENCE A35826  
#authors Rapp, G.; Freudenstein, J.; Klaudiny, J.; Mucha, J.; Wempe, F.; Zimmer, M.; Scheit, K.H.  
#journal DNA Cell Biol. (1990) 9:479-485  
#title Characterization of three abundant mRNAs from human ovarian granulosa cells.

##cross-references MUID:91025550  
#accession D35826

##molecule\_type preliminary  
##residues 1-57 #label RAP  
##cross-references GB:M38188; NID:9189378; PID:9189380  
SUMMARY #length 57 #molecular\_weight 6634 #checksum 8138

Query Match 70.0%; Score 49; DB 2; Length 57;  
Best Local Similarity 75.0%; Pred. No. 2.50e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db: 34 PYPIRRAL 41  
1111111

OY 2 PYPIVRS 9

RESULT 12  
ENTRY S75541 #type complete  
TITLE hypothetical protein sll1218 - *Synechocystis* sp. (PCC 6803)  
ORGANISM #formal\_name *Synechocystis* sp.  
#variety PCC 6803  
DATE 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Sep-1997

## ACCESSIONS

## REFERENCE

#authors

Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naito, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. (1996) 3:109-136  
Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.

#cross-references MUID:97061201

#accession S75541

#status nucleic acid sequence not shown; translation not shown

#molecule\_type DNA

#residues 1-219 #label KAN

#cross-references EMBL:D90911; NID:g1653083; PID:d1018835; PID:g1653186

#note the nucleotide sequence was submitted to the EMBL Data Library, June 1996

## GENETICS

#gene

SUMMARY ycf39 #length 219 #molecular-weight 23534 #checksum 7430

Query Match Best Local Similarity 70.0%; Score 49; DB 2; Length 219;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 143 YPIVRS 150

OY 1 IPYVRS 8

## RESULT 13

ENTRY B64313 #type complete

TITLE hypothetical protein MJ0106 - *Methanococcus jannaschii*ORGANISM #formal\_name *Methanococcus jannaschii*

DATE 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 10-Oct-1997

## ACCESSIONS

#authors

Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.D.; Nguyen, D.; Uitterback, T.R.; Kelley, J.M.; Peterson, K.M.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, H.P.; Fraser, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
Science (1996) 273:1058-1073  
Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.

#accession B64313

#cross-references MUID:96337999

#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule\_type DNA

#residues 1-238 #label BUL

#cross-references GB:U67468; GB:U77117; NID:g1590882; PID:g1590883; TIGR:MJ0106; PID:g1510254

GENETICS #map-position REV102553-101837

SUMMARY #start-codon TTG #length 238 #molecular-weight 26473 #checksum 7921

Query Match Best Local Similarity 70.0%; Score 49; DB 2; Length 238;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 65 IPYVRS 72

OY 1 IPYVRS 8

## RESULT 14

ENTRY MNA6F #type complete

TITLE membrane protein lacF - *Agrobacterium radiobacter*ORGANISM #formal\_name *Agrobacterium radiobacter*

DATE 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 05-Sep-1997

## ACCESSIONS

#authors

#journal

#title

#accession S25248

#molecule\_type DNA

#residues 1-298 #label WIL

#cross-references EMBL:X6596; NID:g38967; PID:g38969

## GENETICS

#gene

CLASSIFICATION lacF

KEYWORDS #superfamily inner membrane protein ugpa

SUMMARY binding protein-dependent transport system; lactose

#length 298 #molecular-weight 33618 #checksum 5973

Query Match Best Local Similarity 70.0%; Score 49; DB 1; Length 298;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 34 YPIVRS 40

OY 3 YPIVRS 9

## RESULT 15

ENTRY C64472 #type complete

TITLE hypothetical protein MJ1380 - *Methanococcus jannaschii*ORGANISM #formal\_name *Methanococcus jannaschii*

DATE 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 10-Oct-1997

## ACCESSIONS

#authors

Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.D.; Nguyen, D.; Uitterback, T.R.; Kelley, J.M.; Peterson, K.M.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, H.P.; Fraser, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
Science (1996) 273:1058-1073  
Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.

#accession C64472

#cross-references MUID:96337999

#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule\_type DNA



US-08-452-843-9.rpr

Page 7

```
#map_position FOR1328821-1329057
SUMMARY #length 78 #molecular-weight 9306 #checksum 2481
```

|                       |        |                     |       |            |
|-----------------------|--------|---------------------|-------|------------|
| Query Match           | 68.68; | Score 48;           | DB 2; | Length 78; |
| Best Local Similarity | 44.48; | Pred. No. 3.78e+01; |       |            |
| Matches               | 4;     | Conservative        | 1;    | Indels 0;  |
|                       |        | Mismatches          | 1;    | Gaps 0;    |

Db 38 TRAVIRAL 46  
QY 1 IPYIVRSI 9

Search completed: Fri Sep 11 13:10:34 1998  
Job time : 26 secs.

**THIS PAGE BLANK (USPTO)**

(W.I.)

Distribution rights by Oxford Molecular Ltd

generated.

1 IPYPIVRSU 9

PAM 150

59111 seqs, 2508

Existing first 45

1:swiss1

mean 24.728; Variance 27.192; scale 0.905

ved by analysis of the total score distribution.

## SUMMARIES

| Accession | Length | DB | ID         | Description             | Pred. No. |
|-----------|--------|----|------------|-------------------------|-----------|
| U00096    | 340    | 1  | RF02_SCHPO | PROBABLE ACTIVATOR 1 4  | 1.86e+01  |
| U00096    | 289    | 1  | RAS_GEOCY  | RAS-LIKE PROTEIN.       | 9.04e+01  |
| U00096    | 407    | 1  | YJGE_ECOLI | HYDROLYTICAL 52.9. KD P | 9.04e+01  |
| U00096    | 342    | 1  | AGC1_ARABN | ARINASE (EC 3.5.3.1).   | 1.51e+00  |
| U00096    | 381    | 1  | Y931_METHA | HYDROLYTICAL PROTEIN M  | 2.51e+00  |
| U00096    | 356    | 1  | OXDA_TRIYR | D-AMINO ACID OXIDASE (  | 4.13e+00  |
| U00096    | 1061   | 1  | RNE_ECOLI  | RIBONUCLEASE E (EC 3.1  | 4.13e+00  |
| U00096    | 1676   | 1  | CO5_HUMAN  | COMPLEMENT C5 PRECURSO  | 4.13e+00  |
| U00096    | 1680   | 1  | CO5_MOUSE  | COMPLEMENT C5 PRECURSO  | 4.13e+00  |
| U00096    | 4344   | 1  | DYHC_MEANI | DYNEIN HEAVY CHAIN, CY  | 4.13e+00  |
| U00096    | 4349   | 1  | DYHC_MECHA | DYNEIN HEAVY CHAIN, CY  | 4.13e+00  |
| U00096    | 4367   | 1  | DYHC_NEUTR | DYNEIN HEAVY CHAIN, CY  | 4.13e+00  |
| U00096    | 238    | 1  | Y106_METHA | HYDROLYTICAL PROTEIN M  | 6.75e+00  |
| U00096    | 298    | 1  | LACF_AGRAD | LACTOSE TRANSPORT SYST  | 6.75e+00  |
| U00096    | 78     | 1  | Y80_METHA  | HYDROLYTICAL PROTEIN M  | 1.10e+01  |
| U00096    | 305    | 1  | LYGD_PSPA  | C ALPHA-DEHYDROGENASE   | 1.10e+01  |
| U00096    | 356    | 1  | YJGC_SCHPO | HYDROLYTICAL 41.3 KD P  | 1.10e+01  |
| U00096    | 440    | 1  | NNM1_YEAST | NAAI PROTEIN PRECURSOR  | 1.10e+01  |
| U00096    | 531    | 1  | TRPC_PHLAR | INDOLE-3-GLYCEROL PHOS  | 1.10e+01  |
| U00096    | 880    | 1  | RAPI_SULAC | DNA-DIRECTED RNA POLYM  | 1.10e+01  |
| U00096    | 268    | 1  | CECE_SYN2D | PHOSPHOCYANOBILIN LYASE | 1.76e+01  |
| U00096    | 272    | 1  | CYNT_SYN2P | CARBONIC ANHYDRASE (EC  | 1.76e+01  |
| U00096    | 315    | 1  | HTRE_HAEIN | LIPID A BIOSYNTHESIS L  | 1.76e+01  |

|    |    |      |      |   |             |                         |          |
|----|----|------|------|---|-------------|-------------------------|----------|
| 25 | 47 | 67.1 | 634  | 1 | HS71_LEIMA  | MITOCHONDRIAL HEAT SHO  | 1.76e+01 |
| 26 | 47 | 67.1 | 675  | 1 | HS7M_PEA    | MITOCHONDRIAL HEAT SHO  | 1.76e+01 |
| 27 | 47 | 67.1 | 1056 | 1 | YNN2_KEAT   | HYPOTHEETICAL 119.3 KD  | 1.76e+01 |
| 28 | 47 | 67.1 | 1082 | 1 | RLB2_HUDAN  | RETINOBLASTOMA-LIKE PR  | 1.76e+01 |
| 29 | 47 | 67.1 | 1230 | 1 | UGS4_SOLTU  | SOLUBLE GLYCOGEN (STAR  | 1.76e+01 |
| 30 | 47 | 67.1 | 1679 | 1 | Y109_YEAST  | HYPOTHEETICAL 195.1 KD  | 1.76e+01 |
| 31 | 46 | 65.7 | 118  | 1 | R121_MEJUA  | 50S RIBOSOMAL PROTEIN   | 2.81e+01 |
| 32 | 46 | 65.7 | 284  | 1 | Y309_MEJUA  | HYPOTHEETICAL PROTEIN M | 2.81e+01 |
| 33 | 46 | 65.7 | 291  | 1 | BACH_NAIPH  | HALORHODOPSIN (HR).     | 2.81e+01 |
| 34 | 46 | 65.7 | 607  | 1 | G6P1_TRYBB  | GLUCOSE-6-PHOSPHATE IS  | 2.81e+01 |
| 35 | 46 | 65.7 | 682  | 1 | HS7M_SOLTU  | MITOCHONDRIAL HEAT SHO  | 2.81e+01 |
| 36 | 46 | 65.7 | 1175 | 1 | SR52_YEAST  | ADP-DEPENDENT DNA HELI  | 2.81e+01 |
| 37 | 46 | 65.7 | 1330 | 1 | VCAP_PRYIS  | MAJOR CAPSID PROTEIN (  | 2.81e+01 |
| 38 | 46 | 65.7 | 1376 | 1 | VCAP_HSYEB  | MAJOR CAPSID PROTEIN (  | 2.81e+01 |
| 39 | 46 | 65.7 | 1441 | 1 | VGIM_BUNVL7 | M POLYPROTEIN PRECURSO  | 2.81e+01 |
| 40 | 46 | 65.7 | 1453 | 1 | VP15_YEAST  | PROTEIN KINASE VP515 (  | 2.81e+01 |
| 41 | 45 | 64.3 | 384  | 1 | Y243_MEJUA  | HYPOTHEETICAL PROTEIN M | 4.43e+01 |
| 42 | 45 | 64.3 | 799  | 1 | YCV2_YEAST  | HYPOTHEETICAL 91.6 KD P | 4.43e+01 |
| 43 | 45 | 64.3 | 824  | 1 | YOT5_CABEL  | HYPOTHEETICAL 95.7 KD P | 4.43e+01 |
| 44 | 45 | 64.3 | 1451 | 1 | A2M2_MOUSE  | MURINGLOBULIN 2 PRECU   | 4.43e+01 |
| 45 | 45 | 64.3 | 1663 | 1 | CO3_MOUSE   | COMPLEMENT C3 PRECURSO  | 4.43e+01 |

## ALIGNMENTS

```

ID      REFCD SCHPO      STANDARD;      PRT;      340 AA.
AC      Q09843;
DT      01-FEB-1996 (REL. 33, CREATED)
DT      01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT      01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE      PROBABLE ACTIVATOR 1 41 KD SUBUNIT (REPLICATION FACTOR C 41 KD
DE      SUBUNIT).
GN      SPAC32D3.02.
OS      SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC      EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-972;
RA      NIBLEY D., HARRIS D., BARRELL B.G., RAUANDREAM M.A., WALSH S.V.;
RL      SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC      -1- FUNCTION: THE ELONGATION OF PRIME DNA TEMPLATES BY DNA POLYMERASE
CC      DELTA AND EPSILON REQUIRES THE ACTION OF THE ACCESSORY PROTEINS
CC      PROLFEERING CELL NUCLEAR ANTIGEN (PCNA) AND ACTIVATOR 1. THE
CC      41 KD SUBUNIT BINDS ATP AND TO SINGLE-STRANDED DNA
CC      (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE ACTIVATOR 1 36 TO 40 KD SUBUNITS
CC      FAMILY.
DR      EMBL: Z64354; E205682; -
KW      HYPOTHETICAL PROTEIN; DNA REPLICATION; ATP-BINDING; NUCLEAR PROTEIN;
FT      DNA-BINDING.
FT      NP-BIND
SQ      SEQUENCE      340 AA; 37876 MW; FBS18443 CRC32;

Query Match      80.0% ; Score 56; DB 1; Length 340;
Best Local Similarity 66.7% ; Pred. No. 1.86e-01;
Matches      6; Conservative      2; Mismatches      1; Indels      0; Gaps
Db      249 VPYNIIRSL 257
Oy      1 IPYIVRSL 9

RESULT      2
AC      P24498;
ID      RAS.GEOCY      STANDARD;      PRT;      209 AA.
DT      01-MAR-1992 (REL. 21, CREATED)
DT      01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT      01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE      RAS-LIKE PROTEIN.
DE      GEODIA CYDONIUM (SPONGE).
OS

```

OC EUKARYOTA; METAZOA; PORIFERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 9106138.  
 RA ROBITZKI A., SCHROEDER H.C., UGAROVIC D., KUCHINO Y., KURELEC B.,  
 RL GAMULIN V., MOELLER W.E.G.;  
 CC EUR. J. BIOCHEM. 192:499-506(1990).  
 CC -1- FUNCTION: THIS PROTEIN IS ACTIVATED BY THE INSULIN/INSULIN  
 (INSULIN-LIKE)-RECEPTOR SYSTEM. THIS TRANSDUCTION ENABLES THE RAS  
 PROTEIN TO INTERACT WITH THE LECTIN-RECEPTOR/LECTIN COMPLEX, A  
 PROCESS WHICH ULTIMATELY LEAD TO AN INITIATION OF AN INTRA-  
 CELLULAR SIGNAL-TRANSDUCTION CHAIN.  
 CC -1- PIV: PHOSPHORYLATED IN THE PRESENCE OF INSULIN.  
 DR EMBL: M30929; E51958; ALT\_SEQ.  
 DR PIR: S13179; S13179.  
 DR HSSP: P01112; 1PL1.  
 KW GTP-BINDING; PRENYLATION; LIPOPROTEIN; PHOSPHORYLATION.  
 FT NP\_BIND 10 17 GTP (BY SIMILARITY).  
 FT NP\_BIND 79 83 GTP (BY SIMILARITY).  
 FT NP\_BIND 140 143 GTP (BY SIMILARITY).  
 FT DOMAIN 55 63 EFFECTOR REGION (BY SIMILARITY).  
 FT MOD\_RES 58 58 PHOSPHORYLATION (POTENTIAL).  
 FT LIPID 206 206 GERANYL-GERANYL (BY SIMILARITY).  
 SO SEQUENCE 209 AA; 23854 MW; E07739EF CRC32;  
 Query Match 75.7%; Score 53; DB 1; Length 209;  
 Best Local Similarity 66.7%; Pred. No. 9.04e-01;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 178 IPYSLVREL 186  
 QY 1 IPYIVRSL 9  
 RESULT 3  
 ID YGJE\_ECOLI STANDARD; PRT; 487 AA.  
 AC P39414; Q46870;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHELTICAL 52.9 K D PROTEIN IN TTDB-RPSU INTERGENIC REGION.  
 GN YGJE.  
 OS ESCHERICHIA COLI.  
 CC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 CC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA BLATTNER F.R., PLUNKETT G. III, MAYHEM G.F., PERNA N.T., GLASNER F.D.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87248073.  
 RA NESIN M., LUPSKI J.R., SVEC P., GODSON G.N.;  
 RL GENE 51:149-161(1987).  
 RN [3]  
 RP IDENTIFICATION.  
 RX MEDLINE: 95075659.  
 RA BORODOVSKY M., RUDD K.E., KOONIN E.V.;  
 RL NUCLEIC ACIDS RES. 22:4756-4767(1994).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 (PROBABILE).  
 CC -1- SIMILARITY: BELONGS TO THE NADC/P/PHO87 FAMILY OF TRANSPORTERS.  
 CC SODIUM SUBFAMILY.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS  
 IN POSITIONS 35, 51, 132, 245, 268 AND 443.  
 CC EMBL: U28379; G882586;  
 CC EMBL: AE000388; G1789444;  
 DR EMBL: M16194; NOT\_ANNOTATED\_CDS.  
 DR ECOGENE: EG1393; YGJE.  
 KW HYPOTHELTICAL PROTEIN; TRANSMEMBRANE; INNER MEMBRANE; TRANSPORT.  
 FT TRANSMEM 10 30 POTENTIAL.  
 FT TRANSMEM 33 53 POTENTIAL.

FT TRANSMEM 54 74 POTENTIAL.  
 FT TRANSMEM 93 113 POTENTIAL.  
 FT TRANSMEM 137 157 POTENTIAL.  
 FT TRANSMEM 189 209 POTENTIAL.  
 FT TRANSMEM 236 256 POTENTIAL.  
 FT TRANSMEM 292 312 POTENTIAL.  
 FT TRANSMEM 313 333 POTENTIAL.  
 FT TRANSMEM 340 360 POTENTIAL.  
 FT TRANSMEM 370 390 POTENTIAL.  
 FT TRANSMEM 393 413 POTENTIAL.  
 FT TRANSMEM 418 438 POTENTIAL.  
 FT TRANSMEM 465 485 POTENTIAL.  
 FT TRANSMEM 404 404 L -> P (IN REF. 2).  
 FT CONFLICT 457 457 A -> T (IN REF. 2).  
 SO SEQUENCE 487 AA; 52906 MW; EB673FE9 CRC32;  
 Query Match 75.7%; Score 53; DB 1; Length 487;  
 Best Local Similarity 66.7%; Pred. No. 9.04e-01;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 165 ITPPIRNL 173  
 QY 1 IPYIVRSL 9  
 RESULT 4  
 ID ARG1\_ARATH STANDARD; PRT; 342 AA.  
 AC P4637;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ARGINASE (EC 3.5.3.1).  
 DE ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OS EUKARYOTA; PLANTA; EMERYPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;  
 CC CAPRARIACE; CRUCIFERAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, LANDSBERG ERECTA;  
 RX MEDLINE: 95288383.  
 RA KROMPELMAN P.M., FREYERMUTH S.K., CANNON J.F., FINK G.R.,  
 RL POLACCO J.C.;  
 RL PLANT PHYSIOL. 107:1479-1480(1995).  
 CC -1- CATALYTIC ACTIVITY: L-ARGININE + H(2)O - L-ORNITHINE + UREA.  
 CC -1- COFACTOR: NM(2+). (BY SIMILARITY).  
 CC -1- PATHWAY: FIRST STEP IN ARGININE DEGRADATION.  
 CC -1- SIMILARITY: BELONGS TO THE ARGINASE FAMILY.  
 DR EMBL: U15019; G602422;  
 DR PROSITE: PS00147; ARGINASE\_1; 1.  
 DR PROSITE: PS00148; ARGINASE\_2; 1.  
 DR PROSITE: PS01053; ARGINASE\_3; 1.  
 KW HYDROLASE; ARGININE METABOLISM; MANGANESE.  
 FT METAL 161 161 MANGANESE 1 (BY SIMILARITY).  
 FT METAL 185 185 MANGANESE 1 AND 2 (BY SIMILARITY).  
 FT METAL 187 187 MANGANESE 2 (BY SIMILARITY).  
 FT METAL 189 189 MANGANESE 1 (BY SIMILARITY).  
 FT METAL 270 270 MANGANESE 1 AND 2 (BY SIMILARITY).  
 FT METAL 272 272 MANGANESE 2 (BY SIMILARITY).  
 SO SEQUENCE 342 AA; 37344 MW; 9640021A CRC32;  
 Query Match 74.3%; Score 52; DB 1; Length 342;  
 Best Local Similarity 55.6%; Pred. No. 1.51e+00;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Db 163 IYPEVRAV 171  
 QY 1 IPYIVRSL 9  
 RESULT 5  
 ID Y931\_METYA STANDARD; PRT; 381 AA.  
 AC Q58341;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL PROTEIN M0931.  
 GN M0931  
 OS METHANOCOCCUS JANNASCHII.  
 OC ARCHAEABACTERIA: EURYARCHAEOTA: METHANOCOCCALES: METHANOCOCCACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96337999.  
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
 RA KERLEVAGE A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,  
 RA OVERBERG R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,  
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,  
 RA KLENN H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.,  
 RL SCIENCE 273:1058-1073(1996).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0008 FAMILY.  
 DR EMBL: U67536; G1591602; -  
 KM TIGR: M0931; -  
 SO SEQUENCE 381 AA; 43436 MW; 853CF1A9 CRC32;  
 Query Match 72.9%; Score 51; DB 1; Length 381;  
 Best Local Similarity 66.7%; Pred. No. 2.51e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 DB 319 INPIRPL 327  
 QY 1 IPIPIVRS 9  
 RESULT 6  
 ID OXDA-TRIVR STANDARD; PRT; 356 AA.  
 AC O99042;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE D-AMINO ACID OXIDASE (EC 1.4.3.3) (DAMOXY) (DAAO).  
 OS TRIGONOPSIS VARIABILIS.  
 OC EUKARYOTA: FUNGI: DEUTEROMYCOTINA (IMPERFECT FUNGI).  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CBS 4095;  
 RA GONZALEZ F.;  
 RL THESES (1996), UNIVERSIDAD DE SALAMANCA, SPAIN.  
 CC -1- CATALYTIC ACTIVITY: A D-AMINO ACID + H(2)O + O(2) -> A 2-OKO-ACID +  
 CC NH(3) + H(2)O(2).  
 CC -1- COFACTOR: FAD FLAVOPROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.  
 DR EMBL: Z50019; E187982; -  
 DR PROSITE: PS00677; DAO; 1.  
 KW OXIDOREDUCTASE: FLAVOPROTEIN; FAD.  
 FT NP\_BIND 4 18 FAD (ADP PART) (POTENTIAL).  
 FT ACT\_SITE 243 243 BY SIMILARITY.  
 FT ACT\_SITE 324 324 BY SIMILARITY.  
 FT SEQUENCE 356 AA; 39301 MW; BA069642 CRC32;  
 Query Match 71.4%; Score 50; DB 1; Length 356;  
 Best Local Similarity 55.6%; Pred. No. 4.13e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 DB 66 VSPILREL 74  
 QY 1 IPIPIVRS 9  
 RESULT 7  
 ID RNE-ECOLI STANDARD; PRT; 1061 AA.  
 AC P21513;  
 DT 01-MAY-1991 (REL. 18, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE RIBONUCLEASE E (EC 3.1.4.-) (RNASE E).  
 GN RNE OR AMS OR HMP1.  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA: GRACILICUTES: SCOTOBACTERIA: FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE OF 1-1025 FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE: 93078265.  
 RA CASAREGOLA S., JACO A., LAUDY D., MCGURK G., MARGARSON S.,  
 RA TEMPLE M., NORRIS V., HOLLAND I.B.;  
 RL J. MOL. BIOL. 228:30-40(1992).  
 RN [2]  
 RP SEQUENCE OF 1-844 FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE: 91131576.  
 RA CLAVIERIE-MARTIN F., DIAZ-TORRES M., YANCEY S.D., KUSHNER S.R.;  
 RL J. BIOL. CHEM. 266:2843-2851(1991).  
 RN [3]  
 RP PARTIAL SEQUENCE FROM N.A., AND SEQUENCE OF 1-27.  
 RC STRAIN-K12;  
 RX MEDLINE: 91187608.  
 RA CHAUDHAN A.K., MICZAK A., TARASEVICIENE L., APIRION D.;  
 RL NUCLEIC ACIDS RES. 19:125-129(1991).  
 RN [4]  
 RP SEQUENCE OF 844-1061 FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN-K12;  
 RX MEDLINE: 94022304.  
 RA CORNACK R.S., GENEVAUX J.L., MACKIE G.A.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 90:9006-9010(1993).  
 CC -1- FUNCTION: THIS PROTEIN MATURES 5S RNA FROM ITS PRECURSORS FROM  
 CC ALL THE RNA GENES. IT ALSO CLEAVES RNA I, A MOLECULE THAT  
 CC CONTROLS THE REPLICATION OF COLEI PLASMID DNA. IT IS THE MAJOR  
 CC ENDOIRONUCLEASE PARTICIPATING IN MRNA TURNOVER IN E. COLI.  
 CC -1- SUBUNIT: ORGANISED INTO A STRUCTURE (PROCESSOME OR RNA  
 CC DEGRADOSOME) CONTAINING A NUMBER OF RNA-PROCESSING ENZYMS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASTIC.  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1003  
 CC ONWARD AND IS SHORTER (1025 AA) DUE TO A FRAMESHIFT.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE C-TERMINUS  
 CC AND IS SHORTER (815 AA) DUE TO A FRAMESHIFT.  
 CC -1- CAUTION: REF.3 SEQUENCE WAS ALSO INCORRECT IN MANY POSITIONS DUE  
 CC TO FRAMESHIFTS.  
 DR EMBL: X67470; G49116; ALT\_FRAME.  
 DR EMBL: M62747; G145273; ALT\_FRAME.  
 DR EMBL: X54309; G42773; ALT\_FRAME.  
 DR EMBL: L23942; G397760; -  
 DR PIR: JG0009; JG0009.  
 DR PIR: A23747; A23747.  
 DR PIR: S25116; S25116.  
 DR PIR: S27311; S27311.  
 DR ECGENE: EC10859; RNE.  
 KW HYDROLASE: NUCLEASE: ENDONUCLEASE: RNA-BINDING.  
 FT CONFLICT 390 390 Q -> H (IN REF. 2).  
 FT CONFLICT 364 364 R -> A (IN REF. 2).  
 FT CONFLICT 784 784 K -> N (IN REF. 2).  
 FT CONFLICT 838 838 P -> R (IN REF. 2).  
 FT CONFLICT 905 905 P -> R (IN REF. 1).  
 FT SEQUENCE 1061 AA; 118301 MW; 347377BC CRC32;  
 Query Match 71.4%; Score 50; DB 1; Length 1061;  
 Best Local Similarity 75.0%; Pred. No. 4.13e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 DB 844 IPIPIVRS 851  
 QY 1 IPIPIVRS 8  
 RESULT 8  
 ID COS-HUMAN STANDARD; PRT; 1676 AA.  
 AC P01031;  
 DT 21-JUL-1986 (REL. 01, CREATED)

01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 COMPLEMENT C5 PRECURSOR (CONTAINS: C5A ANAPHYLATOXIN).  
 CS  
 HOMO (SAPIENS (HUMAN)).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91079575.  
 RA HAVILAND D.L., HAVILAND J.C., FLEISCHER D.T., HUNT A., WETSEL R.A.;  
 RL J. CLIMINOL. 146:362-368(1991).  
 CC [2]  
 RP SEQUENCE OF 412-1676 FROM N.A.  
 RX MEDLINE: 88209511.  
 RA WETSEL R.A., LEMONS R.S., LEBEAU M.M., BARNDON S.R., NOACK D.,  
 RL TACK B.F.;  
 RL BIOCHEMISTRY 27:1474-1482(1988).  
 CC [3]  
 RP SEQUENCE OF 412-902 FROM N.A.  
 RX MEDLINE: 85130937.  
 RA LUNDWALL A.B., WETSEL R.A., KRISTENSEN T., WHITEHEAD A.S.,  
 RL WOODS D.E., OGDEN R.C., COLTEN H.R., TACK B.F.;  
 RL J. BIOL. CHEM. 260:2108-2112(1985).  
 CC [4]  
 RP SEQUENCE OF 678-751.  
 RX MEDLINE: 79005687.  
 RA FERNANDEZ H.N., HUGLI T.E.;  
 RL J. BIOL. CHEM. 253:6955-6964(1978).  
 CC [5]  
 RP SEQUENCE OF 678-751 FROM N.A.  
 RX MEDLINE: 91144547.  
 RA BOHSACK J.F., MOLLISON K.W., BUKO A.M., ASHWORTH J.C., HILL H.R.;  
 RL BIOCHEM. J. 273:695-694(1991).  
 CC [6]  
 RP STRUCTURE BY NMR OF C5A.  
 RX MEDLINE: 88309754.  
 RA ZUIDERWEG E.R., MOLLISON K.W., HENKIN J., CARTER G.W.;  
 RL BIOCHEMISTRY 27:3568-3580(1988).  
 CC [7]  
 RP STRUCTURE BY NMR OF C5A.  
 RX MEDLINE: 89207527.  
 RA ZUIDERWEG E.R., NETTESHEIM D.G., MOLLISON K.W., CARTER G.W.;  
 RL BIOCHEMISTRY 28:172-185(1989).  
 CC [8]  
 RP STRUCTURE BY NMR OF C5A.  
 RX MEDLINE: 89274164.  
 RA ZUIDERWEG E.R., FESIK S.W.;  
 RL BIOCHEMISTRY 28:2387-2391(1989).  
 CC [9]  
 RP STRUCTURE BY NMR OF C5A.  
 RX MEDLINE: 97160477.  
 RA ZHANG X., BOYAR W., GALAKATOS N., GONNELLA N.C.;  
 RL PROTEIN SCI. 6:65-72(1997).  
 CC [10]  
 RP STRUCTURE BY NMR OF C5A.  
 RX MEDLINE: 97332508.  
 RA ZHANG X., BOYAR W., TOTI M.J., WENNOGLE L., GONNELLA N.C.;  
 RL PROTEINS 28:261-267(1997).  
 CC -1- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE  
 SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,  
 INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE  
 FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYTIC  
 COMPLEX IS ASSEMBLED.  
 CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,  
 C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT  
 INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR  
 PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND  
 POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR  
 MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).  
 CC -1- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC  
 RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE

BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN,  
 RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA  
 CHAIN).  
 CC -1- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.  
 CC -1- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.  
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 855  
 ONWARD DUE TO THE PRESENCE OF AN ALU REPEAT.  
 CC EMBL: M57729; G179983; -;  
 DR EMBL: M65134; G179692; -;  
 DR PIR: A40075; C5HD.  
 DR PIR: S15121; S15121.  
 DR PDB: 1KJS; 15-MAY-97.  
 DR PDB: 1CFA; 17-SEP-97.  
 DR MIN: 120900; -;  
 DR PROSITE: PS00477; ALPHA-2-MACROGLOBULIN; FALSE\_NEG.  
 DR PROSITE: PS01178; ANAPHYLATOXIN\_1; 1.  
 DR PROSITE: PS01178; ANAPHYLATOXIN\_2; 1.  
 KW COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; GLYCOPROTEIN;  
 KW PLASMA; MEMBRANE ATTACK COMPLEX; CYTOLYSIS; INFLAMMATORY RESPONSE;  
 KW SIGNAL; POLYMORPHISM; 3D-STRUCTURE.  
 FT SIGNAL 1 18  
 FT CHAIN 1 673 COMPLEMENT C5 BETA CHAIN.  
 FT PROPEP 674 677  
 FT CHAIN 678 1676 COMPLEMENT C5 ALPHA CHAIN.  
 FT PEPTIDE 678 751 C5A ANAPHYLATOXIN.  
 FT CHAIN 752 1676 C5B (ALPHA).  
 FT DOMAIN 698 732 ANAPHYLATOXIN-LIKE.  
 FT DISULFID 698 724  
 FT DISULFID 711 732  
 FT CARBOHYD 741 741  
 FT CARBOHYD 911 911  
 FT CARBOHYD 1115 1115 POTENTIAL.  
 FT CARBOHYD 1630 1630 POTENTIAL.  
 FT VARIANT 518 518 F->S.  
 SQ SEQUENCE 1676 AA; 188331 MW; 9D5C6E59 CRC32;  
 Query Match 71.4%; Score 50; DB 1; Length 1676;  
 Best Local Similarity 62.5%; Pred. No. 4,136+00;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 Db 829 IPYSVVG 836  
 Qy 1 IPYIVRS 8  
 RESULT 9  
 ID C05\_MOUSE STANDARD: PRT; 1680 AA.  
 AC P06684;  
 DT 01-JAN-1988 (REL. 06, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT C5 PRECURSOR (CONTAINS: C5A ANAPHYLATOXIN).  
 GN C5 OR HC.  
 OS MUS MUSCULUS (MOUSE).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; RODENTIA.  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 90153853.  
 RA WETSEL R.A., FLEISCHER D.T., HAVILAND D.L.;  
 RL J. BIOL. CHEM. 265:2435-2440(1990).  
 CC [2]  
 RP SEQUENCE OF 41-1680 FROM N.A.  
 RX MEDLINE: 87185363.  
 RA WETSEL R.A., OGATA R.T., TACK B.F.;  
 RL BIOCHEMISTRY 26:737-743(1987).  
 CC -1- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE  
 SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,  
 INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE  
 FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYTIC  
 COMPLEX IS ASSEMBLED.  
 CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,

CC C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT  
CC INDICES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR  
CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND  
CC BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF  
CC POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR  
CC MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).  
CC -1 SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC  
CC RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE  
CC BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN, ALPHAF  
CC RELASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA  
CC CHAIN).  
CC -1 SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.  
CC -1 SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.  
CC EMBL: M35525; G309124; -  
CC EMBL: M35526; G309123; -  
CC PIR: A27538; A27538.  
CC PIR: A35530; A35530.  
CC HSP: P01032; ICSA.  
CC MGI: 96031; HC.  
CC PROSITE: PS00477; ALPHA\_2\_MACROGLOBULIN; FALSE\_NEG.  
CC PROSITE: PS01177; ANAPHYLATOXIN\_1; 1.  
CC PROSITE: PS01178; ANAPHYLATOXIN\_2; 1.  
CC COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; GLYCOPROTEIN;  
CC PLASMA; MEMBRANE ATTACK COMPLEX; CYTOLYSIS; INFLAMMATORY RESPONSE;  
CC SIGNAL.  
CC SIGNAL. 1 18  
CC CHAIN 19 1680  
CC CHAIN 19 674  
CC PROPEP 675 678  
CC CHAIN 679 1680  
CC PEPTIDE 679 755  
CC CHAIN 756 1680  
CC DOMAIN 702 736  
CC DISULFID 702 728  
CC DISULFID 703 735  
CC DISULFID 715 736  
CC CARBOHYD 427 427  
CC CARBOHYD 915 915  
CC CARBOHYD 1119 1119  
CC CARBOHYD 1633 1633  
CC VARIANT 216 216  
CC VARIANT 217 1680  
CC SEQUENCE 1680 AA; 188877 MW; AA17044B CRC32;  
Query Match 71.4%; Score 50; DB 1; Length 1680;  
Best Local Similarity 62.3%; Pred. No. 4.13e+00;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
DB 833 IPIVYRG 840  
QY 1 IPIVYRS 8  
RESULT 10  
ID DYHC\_EMENTI STANDARD: PRT: 4344 AA.  
AC P45444;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).  
GN NUDA.  
OS EMERICELLA NIDULANS (ASPERGILLUS NIDULANS).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; PLECTOMYCETES; EUROTIALES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94181539.  
RA XIANG X., BECKWITH S.M., MORRIS R.N.;  
RL PROC. NATL. ACADE. SCI. U.S.A. 91:2100-2104(1994).  
CC -1 FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
CC ORGANELLES ALONG MICROTUBULES. REQUIRED TO MAINTAIN UNIFORM  
CC NUCLEAR DISTRIBUTION IN HYPAE.  
CC -1 SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF

CC INTERMEDIATE AND LIGHT CHAINS.  
CC -1 SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1 SIMILARITY: TO OTHER DYNEIN HEAVY CHAINS.  
CC EMBL: U03904; G451539; -  
CC PIR: A53489; A53489.  
CC MOTOR PROTEIN: MICROTUBULES; DYNEIN; ATP-BINDING;  
CC HEPAID REPEAT PATTERN.  
CC NP\_BIND 1933 1940  
CC NP\_BIND 2223 2230  
CC NP\_BIND 2592 2599  
CC NP\_BIND 2932 2939  
CC SEQUENCE 4344 AA; 492470 MW; 1D75C7EB CRC32;  
Query Match 71.4%; Score 50; DB 1; Length 4344;  
Best Local Similarity 75.0%; Pred. No. 4.13e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
DB 394 PYPIRRL 401  
QY 2 PYPIVRS 9  
RESULT 11  
ID DYHC\_NECHE STANDARD: PRT: 4349 AA.  
AC P78716;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).  
GN DCL.  
OS NECTRIA HAMATOCOCOA.  
OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-T213;  
RC INOGE S., AIST J.R., TURGEON B.G., YODER O.C.;  
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBI DATA BANKS.  
CC -1 FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
CC ORGANELLES ALONG MICROTUBULES.  
CC -1 SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF  
CC INTERMEDIATE AND LIGHT CHAINS.  
CC -1 SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1 SIMILARITY: TO OTHER DYNEIN HEAVY CHAINS.  
CC EMBL: U84215; G1814018; -  
CC MOTOR PROTEIN: MICROTUBULES; DYNEIN; ATP-BINDING;  
CC HEPAID REPEAT PATTERN.  
CC NP\_BIND 1946 1953  
CC NP\_BIND 2239 2246  
CC NP\_BIND 2604 2611  
CC NP\_BIND 2946 2953  
CC SEQUENCE 4349 AA; 493453 MW; 961A2CID CRC32;  
Query Match 71.4%; Score 50; DB 1; Length 4349;  
Best Local Similarity 75.0%; Pred. No. 4.13e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
DB 407 PYPIRRL 414  
QY 2 PYPIVRS 9  
RESULT 12  
ID DYHC\_NECHE STANDARD: PRT: 4367 AA.  
AC P45443;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).  
GN RO-1.  
OS NEUROSPORA CRASSA.  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; PYRENOMYCETES; SORDARIALES.  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-74-OR23-1A;  
RX MEDLINE: 95014704.  
RA PLAMANN M., MIKE P.F., TINSLEY J.H., BRUNO K.S.;  
RL J. CELL BIOL. 127:139-149(1994).  
CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
CC ORGANELLES ALONG MICROTUBULES. REQUIRED TO MAINTAIN UNIFORM  
CC NUCLEAR DISTRIBUTION IN HYPAE.  
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF  
CC INTERMEDIATE AND LIGHT CHAINS.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: TO OTHER DYNEIN HEAVY CHAINS.  
DR EMBL: L31504; G473490; -  
KW MOTOR PROTEIN; MICROTUBULES; DYNEIN; ATP-BINDING;  
KM HEPTAD REPEAT PATTERN.  
FT N-BIND 1943 1950 ATP (POTENTIAL).  
FT N-BIND 2240 2247 ATP (POTENTIAL).  
FT N-BIND 2605 2612 ATP (POTENTIAL).  
FT N-BIND 2947 2954 ATP (POTENTIAL).  
SQ SEQUENCE 4367 AA; 495568 MW; B81B5E92 CRC32;  
Query Match 71.4%; Score 50; DB 1; Length 4367;  
Best Local Similarity 75.0%; Pred. No. 4.13e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 404 PYPIRRL 411  
|||  
QY 2 PYPIRSL 9

RESULT 13  
ID Y106.METJA STANDARD; PRT: 238 AA.  
AC 057570;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL PROTEIN MJ0106.  
GN MJ0106  
OS METHANOCOCCUS JANNASCHII.  
OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96337999.  
RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,  
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKIY M.,  
RA KLENN H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;  
RL SCIENCE 273:1058-1073(1996).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: WEAK, TO M.JANNASCHII MJ1210.  
DR EMBL: U67468; G1590883; -  
KW TIGR: MJ0106; -  
KM HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
FT TRANSMEM 19 39 POTENTIAL.  
FT TRANSMEM 79 99 POTENTIAL.  
FT TRANSMEM 141 161 POTENTIAL.  
SQ SEQUENCE 238 AA; 26473 MW; 6443385A CRC32;  
Query Match 70.0%; Score 49; DB 1; Length 238;  
Best Local Similarity 75.0%; Pred. No. 6.75e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 65 IIPYVRA 72  
|||  
QY 1 IIPYVRS 8

RESULT 14

ID LACE AGRD STANDARD; PRT: 298 AA.  
AC P29823;  
DT 01-APR-1993 (REL. 25, CREATED)  
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE LACE TRANSPORT SYSTEM PERMEASE PROTEIN LACE.  
GN LACE.  
OS AGROBACTERIUM RADIOBACTER.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC RHIZOBIAEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-AR50;  
RX MEDLINE: 92334152.  
RA WILLIAMS S.G., GREENWOOD J.A., JONES C.W.;  
RL MOL. MICROBIOL. 6:1755-1768(1992).  
CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM  
CC FOR LACTOSE. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE  
CC SUBSTRATE ACROSS THE MEMBRANE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, INNER MEMBRANE.  
CC -1- INDUCTION: BY LACTOSE AND VARIOUS GALACTOSIDES, AND SUBJECT TO  
CC CATABOLITE REPRESSION BY GLUCOSE, GALACTOSE AND SUCCINATE. IN  
CC STRAIN AR50 THE EXPRESSION OF THE LAC OPERON IS CONSTITUTIVE.  
CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-  
CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE MALFG  
CC SUBFAMILY.  
DR EMBL: X65596; G38969; -  
DR PIR: S22740; MMAGCF.  
DR PIR: S25248; S25248.  
DR PROSITE: PS00402; BPD\_TRANSP\_INN\_MEMBER. 1.  
KW TRANSPORT; SUGAR TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE.  
FT TRANSMEM 17 37 POTENTIAL.  
FT TRANSMEM 77 97 POTENTIAL.  
FT TRANSMEM 112 132 POTENTIAL.  
FT TRANSMEM 151 171 POTENTIAL.  
FT TRANSMEM 208 228 POTENTIAL.  
FT TRANSMEM 269 289 POTENTIAL.  
SQ SEQUENCE 298 AA; 33618 MW; B353A428 CRC32;  
Query Match 70.0%; Score 49; DB 1; Length 298;  
Best Local Similarity 85.7%; Pred. No. 6.75e+00;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 34 YPIRSL 40  
|||  
QY 3 YPIRSL 9

RESULT 15  
ID YD80.METJA STANDARD; PRT: 78 AA.  
AC 058775;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL PROTEIN MJ1380.  
GN MJ1380  
OS METHANOCOCCUS JANNASCHII.  
OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96337999.  
RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,  
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKIY M.,  
RA KLENN H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;  
RL SCIENCE 273:1058-1073(1996).  
CC -1- SIMILARITY: TO M.JANNASCHII MJ0127 AND MJ0434.  
DR EMBL: U67578; G1592026; -  
KW TIGR: MJ1380; -



```

CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: TO OTHER DYNEIN HEAVY CHAINS.
DR EMBL: U03904; G451539; -.
DR PIR: A53489; A53489.
KW MOTOR PROTEIN; MICROTUBULES; DYNEIN; ATP-BINDING;
KW HEPTAD REPEAT PATTERN.
FT NP_BIND 1933 1940 ATP (POTENTIAL).
FT NP_BIND 2223 2230 ATP (POTENTIAL).
FT NP_BIND 2592 2599 ATP (POTENTIAL).
FT NP_BIND 2932 2939 ATP (POTENTIAL).
SQ SEQUENCE 4344 AA; 492470 MW; 1D75C7EB CRC32;

Query Match 71.4%; Score 50; DB 1; Length 4344;
Best Local Similarity 75.0%; Pred. No. 4.13e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 394 PYPIKRAL 401
QY 2 PYPIKRAL 9
|||||:|

RESULT 11
ID DYHC_NECRA STANDARD; PRT; 4349 AA.
AC P78716;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).
GN DHCI.
OS NECTRIA HAEMATOCOCOA.
OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T213;
RA INOUE S., AIST J.R., TURGEON B.G., YODER O.C.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: TO OTHER DYNEIN HEAVY CHAINS.
DR EMBL: U84215; G1814018; -.
KW MOTOR PROTEIN; MICROTUBULES; DYNEIN; ATP-BINDING;
KW HEPTAD REPEAT PATTERN.
FT NP_BIND 1946 1953 ATP (POTENTIAL).
FT NP_BIND 2239 2246 ATP (POTENTIAL).
FT NP_BIND 2604 2611 ATP (POTENTIAL).
FT NP_BIND 2946 2953 ATP (POTENTIAL).
SQ SEQUENCE 4349 AA; 493453 MW; 961A2C1D CRC32;

Query Match 71.4%; Score 50; DB 1; Length 4349;
Best Local Similarity 75.0%; Pred. No. 4.13e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 407 PYPIKRAL 414
QY 2 PYPIKRAL 9
|||||:|

RESULT 12
ID DYHC_NECRC STANDARD; PRT; 4367 AA.
AC P45443;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).
GN RO-1.
OS NEUROSPORA CRASSA.
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; PYRENOMYCETES; SORDARIALES.
RN [1]

```

RP SEQUENCE FROM N.A.  
RC STRAIN-74-OR3-1A.  
RX MEDLINE; 95014704.  
RA PLAMANN M., MINKE P.F., TINSLEY J.H., BRUNO K.S.;  
RL J. CELL BIOL. 127:139-149(1994).  
CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
CC ORGANELLES ALONG MICROTUBULES. REQUIRED TO MAINTAIN UNIFORM  
CC NUCLEAR DISTRIBUTION IN HYPAE.  
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF  
CC INTERMEDIATE AND LIGHT CHAINS.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: TO OTHER DYNEIN HEAVY CHAINS.  
DR EMBL; L31504; G473480; -  
KW MOTOR PROTEIN; MICROTUBULES; DYNEIN; ATP-BINDING;  
KW HEPTA REPEAT PATTERN.  
FT NP-BIND 1943 1950 ATP (POTENTIAL).  
FT NP-BIND 2240 2247 ATP (POTENTIAL).  
FT NP-BIND 2605 2612 ATP (POTENTIAL).  
FT NP-BIND 2947 2954 ATP (POTENTIAL).  
SO SEQUENCE 4367 AA; 495568 MW; B81B5E92 CRC32;  
Query Match 71.4%; Score 50; DB 1; Length 4367;  
Best Local Similarity 75.0%; Pred. No. 4.13e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 404 PYPIRRL 411  
QY 2 PYPIRSL 9  
RESULT 13  
ID Y106.METUA STANDARD; PRT; 238 AA.  
AC 057570;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL PROTEIN M0106.  
GN M0106.  
OS METHANOCOCCLUS JANNASCHII.  
OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOCOCCEALES; METHANOCOCCEACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96337999.  
RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,  
RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,  
RA SCOTT J.L., GEOCHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORDOVSKI M.,  
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;  
RL SCIENCE 273:1058-1073(1996).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: WEAK, TO M.JANNASCHII M01210.  
DR EMBL; U67468; G1590883;  
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
KW TRANSMEM 19 39 POTENTIAL.  
FT TRANSMEM 79 99 POTENTIAL.  
FT TRANSMEM 141 161 POTENTIAL.  
SO SEQUENCE 238 AA; 26473 MW; 6443385A CRC32;  
Query Match 70.0%; Score 49; DB 1; Length 238;  
Best Local Similarity 75.0%; Pred. No. 6.75e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 65 IYPYVRA 72  
QY 1 IYPYVRS 8  
RESULT 14

ID LAC-AGRD STANDARD; PRT; 298 AA.  
AC P28823;  
DT 01-APR-1993 (REL. 25, CREATED)  
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE LACTOSE TRANSPORT SYSTEM PERMEASE PROTEIN LACF.  
GN LACF.  
OS AGROBACTERIUM RADIIABACTER.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC RHIZOBIAEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AR50.  
RX MEDLINE; 92334152.  
RA WILLIAMS S.G., GREENWOOD J.A., JONES C.W.;  
RL MOL. MICROBIOL. 6:1755-1768(1992).  
CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM  
CC FOR LACTOSE. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE  
CC SUBSTRATE ACROSS THE MEMBRANE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
CC -1- INDUCTION: BY LACTOSE AND VARIOUS GALACTOSIDES, AND SUBJECT TO  
CC CATABOLITE REPRESSION BY GLUCOSE, GALACTOSE AND SUCCINATE. IN  
CC STRAIN AR50 THE EXPRESSION OF THE LAC OPERON IS CONSTITUTIVE.  
CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-  
CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE MALFG  
CC SUBFAMILY.  
DR EMBL; X66596; G38969; -  
DR PIR; S22740; MNA6F.  
DR PIR; S25248; S25248.  
DR PROSITE; PS00402; BPD\_TRANSP\_INN\_MEMB; 1.  
KW TRANSPORT; SUGAR TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE.  
FT TRANSMEM 17 37 POTENTIAL.  
FT TRANSMEM 77 97 POTENTIAL.  
FT TRANSMEM 112 132 POTENTIAL.  
FT TRANSMEM 151 171 POTENTIAL.  
FT TRANSMEM 208 228 POTENTIAL.  
FT TRANSMEM 269 289 POTENTIAL.  
SO SEQUENCE 298 AA; 33618 MW; B353A428 CRC32;  
Query Match 70.0%; Score 49; DB 1; Length 298;  
Best Local Similarity 85.7%; Pred. No. 6.75e+00;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 34 YPIRSL 40  
QY 3 YPIRSL 9  
RESULT 15  
ID YD80.METUA STANDARD; PRT; 78 AA.  
AC 058775;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL PROTEIN M0180.  
GN M0180.  
OS METHANOCOCCLUS JANNASCHII.  
OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOCOCCEALES; METHANOCOCCEACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96337999.  
RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,  
RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,  
RA SCOTT J.L., GEOCHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORDOVSKI M.,  
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;  
RL SCIENCE 273:1058-1073(1996).  
CC -1- SIMILARITY: TO M.JANNASCHII M0127 AND M0434.  
DR EMBL; U67578; G1592026;  
DR TIGR; M01380;

KW HYPOTHETICAL PROTEIN.  
SQ. SEQUENCE 78 AA; 9306 MW; B3785214 CRC32;

Query Match 68.68; Score 48; DB 1; Length 78;  
Best Local Similarity 44.48; Pred. No. 1.10e+01;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```
Db      38 IRYAVIRAL 46
QY      1 IPYPVIRSL 9
```

Search completed: Fri Sep 11 13:10:59 1998  
Job time : 7 secs.

**THIS PAGE BLANK (USPTO)**

Release 3.1A John F. Collins, Biocomputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.

```
Run on:      Fri Sep 11 13:11:17 1998;  MasPar time 3.67 Seconds
```

Title: >US-08-452-843-9  
Description: (1-9) from US08452843.pdf

Scoring table: PAM 150

Searched: 140555 seqs, 42109429 residues

```
Post-processing: Minimum Match 0%
Listing first 45 summaries
```

Database:

```

sptrembl6
1:sp_fungi 2:sp_human 3:sp_invertebrate 4:sp_mammal
5:sp_mhc 6:sp_origanelle 7:sp_phage 8:sp_plant
9:sp_bacteria 10:sp_rodent 11:sp_virus 12:sp_vertbrate
13:sp_unclassified

```

Statistics: Mean 23.989; Variance 30.277; scale 0.7922

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

| Result No. | Score | Query Match | Length | DB | ID     | Description             | Pred. No. |
|------------|-------|-------------|--------|----|--------|-------------------------|-----------|
| 1          | 52    | 74.3        | 133    | 10 | 009002 | TCA.                    | 5.70e+00  |
| 2          | 52    | 74.3        | 133    | 10 | 009006 | SMALL, INDUCIBLE CYTOKI | 5.70e+00  |
| 3          | 51    | 72.9        | 112    | 11 | 084619 | GENOME, PARTIAL SEQUEN  | 9.00e+00  |
| 4          | 51    | 72.9        | 398    | 9  | 007717 | ACEAB.                  | 9.00e+00  |
| 5          | 51    | 72.9        | 410    | 9  | 033331 | DEHYDROGENASE.          | 9.00e+00  |
| 6          | 51    | 72.9        | 478    | 9  | 034726 | YLF5 PROTEIN.           | 9.00e+00  |
| 7          | 50    | 71.4        | 486    | 7  | 021944 | GLUCOSYL TRANSFERASE II | 1.41e+01  |
| 8          | 50    | 71.4        | 537    | 9  | 046977 | RNASE E (FRAGMENT).     | 1.41e+01  |
| 9          | 50    | 71.4        | 1061   | 9  | P75591 | RIBONUCLEASE E (EC 3.1  | 1.41e+01  |
| 10         | 49    | 70.0        | 57     | 2  | Q14931 | PROTEIN OF UNKNOWN FUN  | 2.20e+01  |
| 11         | 49    | 70.0        | 219    | 9  | P07429 | HYDROLYTIC 23.5 KD P    | 2.20e+01  |
| 12         | 48    | 68.6        | 349    | 1  | 005430 | CHROMOSOME IV LEFT ARM  | 3.40e+01  |
| 13         | 48    | 68.6        | 432    | 9  | 034978 | YTFP.                   | 3.40e+01  |
| 14         | 48    | 68.6        | 451    | 9  | P30458 | HYDROLYTICAL 50.4 KD P  | 3.40e+01  |
| 15         | 48    | 68.6        | 535    | 3  | Q18862 | F28C6.2.                | 3.40e+01  |
| 16         | 48    | 68.6        | 905    | 3  | 023596 | ZK792.1 (FRAGMENT).     | 3.40e+01  |
| 17         | 48    | 68.6        | 1026   | 1  | 006315 | CHROMOSOME XII COSMID   | 5.23e+01  |
| 18         | 47    | 67.1        | 129    | 11 | 036882 | MA-P17 (FRAGMENT).      | 5.23e+01  |
| 19         | 47    | 67.1        | 132    | 11 | 036877 | MA-P17 (FRAGMENT).      | 5.23e+01  |
| 20         | 47    | 67.1        | 228    | 9  | P95967 | ORF C04028.             | 5.23e+01  |

|    |    |      |      |    |        |                        |           |
|----|----|------|------|----|--------|------------------------|-----------|
| 21 | 47 | 67.1 | 230  | 8  | Q38732 | DAG PROTEIN PRECURSOR. | 5.23e+01  |
| 22 | 47 | 67.1 | 302  | 1  | D12105 | CHROMOSOME XV READING  | 5.23e+01  |
| 23 | 47 | 67.1 | 311  | 9  | Q4805  | ACT1 TRANSFERASE.      | 5.23e+01  |
| 24 | 47 | 67.1 | 319  | 2  | O14626 | PLATELET ACTIVATING RE | 5.23e+01  |
| 25 | 47 | 67.1 | 364  | 9  | P73423 | CEBD PROTEIN.          | 5.23e+01  |
| 26 | 47 | 67.1 | 434  | 9  | O05241 | HYPOHETHELAL 49.5 KD P | 5.23e+01  |
| 27 | 47 | 67.1 | 463  | 11 | D72601 | GAG POLYPROTEIN (FRAG  | 5.23e+01) |
| 28 | 47 | 67.1 | 588  | 8  | P93341 | PHOSPHOLIPID-DE-SPECI  | 5.23e+01) |
| 29 | 47 | 67.1 | 718  | 3  | Q22949 | SEMILAR TO GLYCEROL-3- | 5.23e+01) |
| 30 | 47 | 67.1 | 1135 | 10 | O64700 | RETINOBLASTOMA-LIKE 2  | 5.23e+01) |
| 31 | 47 | 67.1 | 1139 | 2  | O16084 | P130.                  | 5.23e+01) |
| 32 | 47 | 67.1 | 1139 | 2  | O15073 | 130K PROTEIN.          | 5.23e+01) |
| 33 | 46 | 65.7 | 129  | 11 | O80649 | GAG PROTEIN (FRAGMENT) | 7.97e+01) |
| 34 | 46 | 65.7 | 329  | 3  | P19560 | HYPOHETHELAL PROTEIN C | 7.97e+01) |
| 35 | 46 | 65.7 | 331  | 11 | O69581 | HEPESVIRUS TYPE 6 DNA  | 7.97e+01) |
| 36 | 46 | 65.7 | 336  | 9  | O25428 | CONSERVED HYPOHETHEL   | 7.97e+01) |
| 37 | 46 | 65.7 | 339  | 2  | O00398 | PUTATIVE PURINERGIC RE | 7.97e+01) |
| 38 | 46 | 65.7 | 423  | 9  | P73458 | CARBOXYL-TERMINAL PROT | 7.97e+01) |
| 39 | 46 | 65.7 | 500  | 11 | D72629 | GAG POLYPROTEIN (FRAG  | 7.97e+01) |
| 40 | 46 | 65.7 | 524  | 8  | Q38932 | LYCOPENE EPSILON CYCLA | 7.97e+01) |
| 41 | 46 | 65.7 | 586  | 8  | O04335 | HYPOHETHELAL 66.7 KD P | 7.97e+01) |
| 42 | 46 | 65.7 | 996  | 11 | O36638 | DNA-DEPENDENT DNA POLY | 7.97e+01) |
| 43 | 46 | 65.7 | 1330 | 10 | O04534 | MAJOR CASID PROTEIN (  | 7.97e+01) |
| 44 | 46 | 65.7 | 1375 | 11 | Q39283 | COUNTERPART OF HSV-1 G | 7.97e+01) |
| 45 | 46 | 65.7 | 1873 | 11 | O83044 | METHYLTRANSFERASE.     | 7.97e+01) |

## ALIGNMENTS

```

RESULT      1
ID 009002. PRELIMINARY; PRT; 133 AA.
AC 009002.

DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE TCA4.

OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA, METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUROTHERIA, ROSENTIA.

RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-THYMUS.
RA TANABE S., LU Z., LUO Y., QUACKENBUSH E.J., BERMAN M.A.,
RA COLLINS-RACIE L.A., MI S., REILLY C., LO D., JACOBS K.A., DORF M.E.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE; 97400322.
RA HEDRICK J.A., ZLOTNIK A.;
RL J. IMMUNOL. 159:1589-1593(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA HEDRICK J.A., ZLOTNIK A.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF006637; G2209189; -
DR EMBL; AF001980; G2624927; -
SQ SEQUENCE 133 AA; 14558 MW; C0532523 CRC32;

Query Match 74.3%; Score 52; DB 10; Length 133;
Best Local Similarity 75.0%; Pred. No. 5.70e+00;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0.

DB 40 IPYSIVRG 47
OY 1 IPIPIVRS 8

RESULT      2
ID 009006. PRELIMINARY; PRT; 133 AA.
AC 009006.

DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

```

DE SMALL INDUCIBLE CYTOKINE A21 (BETA CHEMOKINE EXODUS-2).  
 GN SCV421.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUDAKIOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; RODENTIA.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE-TOTAL FETUS;  
 RA HROMAS R.A., GRAY P., KLEMSZ M., FIFE K., BROCKMEYER H.;  
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U88322; G2196924; -  
 MD: MGI:1097677; SCV421.  
 SO SEQUENCE 133 AA; 14615 MW; FE1FA5B6 CRC32;

Query Match 74.3%; Score 52; DB 10; Length 133;  
 Best Local Similarity 75.0%; Pred. No. 5.70e+00;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 40 IPISTVRC 47  
 |||:|:|:  
 QY 1 IPIPIVRS 8

RESULT 3  
 ID Q84619 PRELIMINARY; PRT; 112 AA.  
 AC Q84619;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE GENOME, PARTIAL SEQUENCE.  
 GN A303L.  
 OS PARAMECIUM BURSARIA CHLORELLA VIRUS 1 (PRCV-1).  
 OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PHYCORNNAVIRIDAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 95133167.  
 RA LU Z., LI Y., ZHANG Y., KUTISH G.F., ROCK D.L., VAN ETTEN J.L.;  
 RL VIROLOGY 206:339-352(1995).  
 DR EMBL: U42580; G1181466; -  
 SO SEQUENCE 112 AA; 13416 MW; 5C07006C CRC32;

Query Match  
 Best Local Similarity 72.9%; Score 51; DB 11; Length 112;  
 Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 20 VPESTIRNL 28  
 |||:|:|:  
 QY 1 IPIPIVRS 9

RESULT 4  
 ID Q07717 PRELIMINARY; PRT; 398 AA.  
 AC Q07717;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
 DE ACBAB.  
 GN ACBAB.  
 OS MYCOBACTERIUM TUBERCULOSIS.  
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA OLIVER K., HARRIS D.;  
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA BARRELL B.G., RAJANDREAM M.A., PARKHILL J.;  
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA STRAIN-H37RV;

RX MEDLINE: 96181548.  
 RA PHILIP W.J., POULET S., EIGMEIER K., PASCOPELLA L.;  
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
 RA COLE S.T.;  
 RA PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).  
 DR EMBL: 297193; E324824; -  
 SO SEQUENCE 398 AA; 44582 MW; 95F2E718 CRC32;

Query Match 72.9%; Score 51; DB 9; Length 398;  
 Best Local Similarity 66.7%; Pred. No. 9.00e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 71 IPIYAKSL 79  
 |||:|:|:  
 QY 1 IPIPIVRS 9

RESULT 5  
 ID Q33331 PRELIMINARY; PRT; 410 AA.  
 AC Q33331;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE DEHYDROGENASE.  
 GN MYV002.54C.  
 OS MYCOBACTERIUM TUBERCULOSIS.  
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA MURPHY L., HARRIS D.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE; 96181548.  
 RA PHILIP W.J., POULET S., EIGMEIER K., PASCOPELLA L.;  
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
 RA COLE S.T.;  
 RA PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).  
 DR EMBL: AL008967; E1173919; -  
 SO SEQUENCE 410 AA; 44743 MW; 9BA843CB CRC32;

Query Match  
 Best Local Similarity 72.9%; Score 51; DB 9; Length 410;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 37 PYPIARKL 44  
 |||:|:|:  
 QY 2 PYPIVRS 9

RESULT 6  
 ID Q34726 PRELIMINARY; PRT; 478 AA.  
 AC Q34726;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE YFELS PROTEIN;  
 GN YFELS.  
 OS BACILLUS SUBTILIS.  
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,  
 RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,  
 RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,

RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,  
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,  
 RA DENTON F., DEVINE K.M., DUSTERHOFF A., EHRLICH S.D., EMERSON P.T.,  
 RA ENRIAN K.D., ERLINGTON J., FARET C., FERRARI E., FOUGER D.,  
 RA FRITZ C., FUJITA M., FUJITA Y., FOWA S., GALIZZI A., GALLERON N.,  
 RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,  
 RA GISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,  
 RA HILBERT H., HOLSAPEL S., HOSONO S., HULLO M.F., IYAYA M., JONES L.,  
 RA JORIS B., KARAWATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,  
 RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,  
 RA KORITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,  
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,  
 RA MEDINA N., MELIADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,  
 RA NOONE D., O'REILLY M., OGAWA K., OGIMARA A., ODEGA B., PARK S.H.,  
 RA PARO V., POHL T.M., PORTERELLE D., PORNOULIK S., PRESCOTT A.M.,  
 RA PRESCAN E., PUJIC P., FURNELLE B., RAPPORT G., REY M., REYNOLDS S.,  
 RA RIEBER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADATE Y.,  
 RA SATO T., SCANTLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,  
 RA SEKIGUCHI J., SEKOSKA A., SERO S.J., SERRO P., SHIN B.S., SOLDO B.,  
 RA SOROKIN A., TACCORI E., TAKAGI T., TAKAHASHI H., TAKEHARA K.,  
 RA TAKUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,  
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,  
 RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,  
 RA WINERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUNOTO K., YATA K.,  
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,  
 RA NATURE 390:249-256(1997).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168.  
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.,  
 RA SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AC327.  
 RA YAMAMOTO H., UCHIYAMA S., NUGROHO F.A., SEKIGUCHI J.,  
 RA GENE 194:191-199(1997).  
 RL EMBL: 299108; E1182747; -  
 DR EMBL: D86417; D1023175; -  
 SO SEQUENCE 478 AA; 51432 MW; 90C9082D CRC32;

Query Match 72.9%; Score 51; DB 9; Length 478;  
 Best Local Similarity 66.7%; Pred. No. 9.00e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 155 IFFPIRSL 163  
 QY 1 IPIPIVRS 9

RESULT 7  
 ID 021944 PRELIMINARY; PRT: 486 AA.  
 AC 021944.  
 DT 01-FEB-1998 (TREMBLREL. 05, CREATED)  
 DT 01-FEB-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DE GINCOOYL TRANSFERASE II.  
 GN GTRI.  
 OS BACTERIOPHAGE SF11.  
 OC VIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MAYRIS M., MANNING P.A., MORONA R.,  
 RL MOL. MICROBIOL. 0:0-0(1997).  
 DR EMBL: AF021347; G2465416; -  
 SO SEQUENCE 486 AA; 55778 MW; 64142E04 CRC32;

Query Match 71.4%; Score 50; DB 7; Length 486;  
 Best Local Similarity 85.7%; Pred. No. 1.41e+01;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 449 YPIVRS 455  
 QY 3 YPIVRS 9

RESULT 8  
 ID 046977 PRELIMINARY; PRT: 537 AA.  
 AC 046977.  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE RNASE E (FRAGMENT).  
 GN RNE.  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-W3110;  
 RX MEDLINE: 93078265.  
 RA CASAREGOLA S., JACO A., LAOUDY D., MCGURK G., MARGARSON S.,  
 RA TEMPEETE M., NORRIS V., HOLLAND I.B.,  
 RA J. MOL. BIOL. 228:30-40(1992).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-W3110;  
 RX MEDLINE: 94238701.  
 RA CASAREGOLA S., JACO A., LAOUDY D., MCGURK G., MARGARSON S.,  
 RA TEMPEETE M., NORRIS V., HOLLAND I.B.,  
 RA J. MOL. BIOL. 238:867-867(1994).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-W3110;  
 RA KIDO M., YAMAMURA K., MITANI T., NIKI H., OGURA T., HIRAGA S.,  
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: 049647; G1236531; -  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SO SEQUENCE 537 AA; 59794 MW; 0CC5E9FC CRC32;

Query Match 71.4%; Score 50; DB 9; Length 537;  
 Best Local Similarity 75.0%; Pred. No. 1.41e+01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 347 IRYPIVR 354  
 QY 1 IPIPIVRS 8

RESULT 9  
 ID P77591 PRELIMINARY; PRT: 1061 AA.  
 AC P77591.  
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
 DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
 DE RIBONUCLEASE E (EC 3.1.4.).  
 GN RNE.  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K-12;  
 RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.,  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA OSHIMA T., AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A.,  
 RA IKEMOTO K., INADA T., ITOH T., KAJIHARA M., KANAI K., KASHIMOTO K.,  
 RA KIMURA S., KITAGAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K.,  
 RA MOBI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., SAITO N.,  
 RA SANEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y.,  
 RA YANO M., HORIUCHI T.,  
 RL DNA RES. 3:137-155(1996).  
 DR EMBL: AB000209; G1787325; -

DR EMBL: D90744; G1651530;  
 KM HYDROLASE.  
 SO SEQUENCE 1061 AA; 118196 MW; AEE3417A CRC32;

Query Match. 71.4%; Score 50; DB 9; Length 1061;  
 Best Local Similarity 75.0%; Pred. No. 1.41e+01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 844 IRYPIVRP 851

OY 1 IRYPIVRS 8

RESULT 10

ID 014991 PRELIMINARY; PRT; 57 AA.

AC 014991

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DE PROTEIN OF UNKNOWN FUNCTION.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAEOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

EN EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-OVARY;

RX MEDLINE: 91025550.

RA RAPP G., FREUDENSTEIN J., KLAUDINY J., MUCHA J., WEMPE F.,

RA ZIMMER M., SCHEIT K.H.,

RL DNA CELL BIOL. 9:479-485(1990).

DR EMBL: M8188; G189380;

SO SEQUENCE 57 AA; 6834 MW; 9F00B7D3 CRC32;

Query Match. 70.0%; Score 49; DB 2; Length 57;  
 Best Local Similarity 75.0%; Pred. No. 2.20e+01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 34 PYPGRSM 41

OY 2 PYPVRS 9

RESULT 11

ID P74029 PRELIMINARY; PRT; 219 AA.

AC P74029

DT 01-FEB-1997 (TREMBLREL. 02, CREATED)

DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)

DE HYPOTHETICAL 23.5 KD PROTEIN.

GN YCF39.

OS SYNECHOCYSTIS SP.

OC EUBACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PCC6803;

RA TABATA S.;

RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-PCC6803;

RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,

RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,

RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARIO K.,

RA OKUMURA S., SHIMO S., TAKEICHI C., WADA T., WATANABE A.,

RA YAMADA M., YASUDA M., TABATA S.;

RL DNA RES. 3:109-136(1996).

DR EMBL: D90911; G1653186;

KM HYPOTHETICAL PROTEIN.

SO SEQUENCE 219 AA; 23534 MW; 37A9EA9C CRC32;

Query Match. 70.0%; Score 49; DB 9; Length 219;  
 Best Local Similarity 62.5%; Pred. No. 2.20e+01;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 143 VPYIVRSP 150

OY 1 VPYIVRS 8

RESULT 12

ID 005430 PRELIMINARY; PRT; 349 AA.

AC 005430

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DE CHROMOSOME IV LEFT ARM (EU) DNA SEGMENT (36687 BP).

GN NML.

OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCOMYCETES.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ALPHA 528C;

RX MEDLINE: 97197972.

RA SAREN A.M., LAANEN P., LEJARCEGUI J.B., PAULIN L.;

RL YEAST 13:65-71(1997).

DR EMBL: Z71781; E237289;

SO SEQUENCE 349 AA; 40818 MW; D0A6BE9E CRC32;

Query Match. 68.6%; Score 48; DB 1; Length 349;  
 Best Local Similarity 66.7%; Pred. No. 3.40e+01;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 284 LPYIVRSL 292

OY 1 LPYIVRS 9

RESULT 13

ID 034978 PRELIMINARY; PRT; 432 AA.

AC 034978

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DE YIP.

GN YIP.

OS BACILLUS SUBTILIS.

OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-168;

RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONTI G.,

RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BOCCHERT S.,

RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRANGELL S.C., BRON S.,

RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANI V., CARTER N.M.,

RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,

RA DENIOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,

RA ERTIAN K.D., ERINGTON J., FABBET C., FERRARI E., FOUGER D.,

RA FRITZ C., FUJITA M., FUJITA Y., FDMA S., GALIZZI A., GALLERON N.,

RA GIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,

RA GUISEPP I.G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,

RA HILBERT H., HOLAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,

RA JORIS B., KARAWATA D., KASAHARA Y., KLAERER-BLANCHARD M., KLEIN C.,

RA KOBAYASHI Y., KOETTER P., KONINGSSTEIN G., KROGH S., KUMANO M.,

RA KORITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAJAREVIC V.,

RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,

RA MEDINA N., MELIADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,

RA NOONE D., O'REILLY M., OGAWA K., OGIMARA A., OUDEGA B., PARK S.H.,

RA PARRO V., POHL T.M., PORTETELLE D., PORMOLIK S., PRESCOTT A.M.,

RA PRESCAN E., PUTIC P., PURNELLE B., RAPPOPORT G., REY M., REYNOLDS S.,

RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADATE Y.,

RA SATO T., SCANLAN L., SCHLEICH S., SCHROETER R., SCOPFONE F.,

RA SEKICUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDO B.,



RA SOROKIN A., TACCONTI E., TAKAGI T., TAKAHASHI H., TAKEMATSU K.,  
RA TAKEUCHI M., TANAKASHI A., TANAKA T., TERPSTRA P., TOGNONI A.,  
RA JOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTI A.,  
RA VIARI A., WABUITS R., WEDLER E., WEDLER H., WEITENEGGER T.,  
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,  
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,  
RU NATURE 390:249-256(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168:  
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.,  
RU SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF008220; G2293313;  
DE EMBL: 299119; E1185872;  
SQ SEQUENCE 432 AA; 45420 MW; ED4765AE CRC32;

Query Match 68.68; Score 48; DB 9; Length 432;  
Best Local Similarity 62.58; Pred. No. 3.40e+01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 399 IXPYIVKA 406  
QY 1 IXPYIVRS 8

RESULT 14  
ID P74054 PRELIMINARY; PRT; 451 AA.  
AC P74054:  
DT 01-FEB-1997 (TREMBLER, 02, CREATED)  
DT 01-FEB-1997 (TREMBLER, 02, LAST SEQUENCE UPDATE)  
DE 01-FEB-1997 (TREMBLER, 02, LAST ANNOTATION UPDATE)  
OS SYNECHOCYSTIS SP.  
OC EUBACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PCC6803;  
RA TABATA S.;  
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PCC6803;  
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K.,  
RA OKUMURA S., SHIMO S., TAKEUCHI C., WADA T., WATANABE A.,  
RA YAMADA M., YASUDA M., TABATA S.;  
RL DNA RES. 3:109-136(1996).  
DR EMBL: D90911; G1653214;  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 451 AA; 50417 MW; 42DCF091 CRC32;

Query Match 68.68; Score 48; DB 9; Length 451;  
Best Local Similarity 71.48; Pred. No. 3.40e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 110 LXPYIVR 116  
QY 1 IXPYIVR 7

RESULT 15  
ID 019862 PRELIMINARY; PRT; 535 AA.  
AC 019862:  
DT 01-NOV-1996 (TREMBLER, 01, CREATED)  
DT 01-NOV-1996 (TREMBLER, 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLER, 01, LAST ANNOTATION UPDATE)  
DE F28C6.2.  
OS CAENORHABDITIS ELIGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BURTON J.;

RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,  
RA COULSON A., CRAXTON M., DEAR S., DG Z., DURBIN R., FAVELLO A.,  
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
RA LATREILLE P., LIGHTNING J., LLOYD C., MCKERRAY A., MORTIMORE B.,  
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIKKEN L., ROOPRA A.,  
RA SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONNHAMMER E.,  
RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 368:32-38(1994).  
DR EMBL: Z68315; E214854;  
SQ SEQUENCE 535 AA; 60877 MW; CD2FB0CD CRC32;

Query Match 68.68; Score 48; DB 3; Length 535;  
Best Local Similarity 66.78; Pred. No. 3.40e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 210 IXPYIVRSL 218  
QY 1 IXPYIVRSL 9

Search completed: Fri Sep 11 13:11:39 1998  
Job time : 22 secs.

**THIS PAGE BLANK (USPTO)**

七 十 五 十 七  
 (23)

Release 3.1A John F. Collins, Biocomputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Fri Sep 11 13:05:44 1998;  MasPar time 2.65 Seconds
            54 863 w111400 001 wdata00000
```

Tabular output not generated.

Title: >US-08-452-843-8  
(1-9) from US084

Perfect Score: 73  
Sequence: 1 TTTTCTTTT 9

Sequence: 1 IPYPIVRKL 9

Scoring table: PAM 150

Gap

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0% Listing first 45 summaries

Database:

1:part1 2:part2 3:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 17.379; Variance 46.997; scale 0.370

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| Query No. | Score | Match Length | DB      | ID     | Description             | Pred. No |
|-----------|-------|--------------|---------|--------|-------------------------|----------|
| 1         | 73    | 100.0        | 9 18    | R89369 | Cw6 consensus peptide   | 7.14e-02 |
| 2         | 65    | 89.0         | 9 18    | R89370 | Cw6 consensus peptide   | 7.54e-01 |
| 3         | 53    | 72.6         | 9 18    | R89371 | Cw6 consensus peptide   | 2.33e-01 |
| 4         | 51    | 69.9         | 123 25  | W28511 | Product of clone L105   | 3.85e-01 |
| 5         | 50    | 68.5         | 356 2   | P70388 | D-aminic acid oxidase   | 5.04e-01 |
| 6         | 50    | 68.5         | 356 3   | R04066 | T-variabils D-amino     | 5.04e-01 |
| 7         | 49    | 67.1         | 1676 15 | W47604 | Pro-C5 polypeptide      | 6.58e-01 |
| 8         | 48    | 65.8         | 159 27  | R44125 | Streptococcus pneumonia | 8.58e-01 |
| 9         | 48    | 65.8         | 311 25  | W32084 | Streptococcus pneumonia | 8.58e-01 |
| 10        | 48    | 65.8         | 833 27  | K32114 | Streptococcus pneumonia | 8.58e-01 |
| 11        | 46    | 63.0         | 245 20  | W08080 | Bovine Oncostatin M     | 1.45e-02 |
| 12        | 46    | 63.0         | 356 29  | W44296 | Rat vascular endothel   | 1.45e-02 |
| 13        | 46    | 63.0         | 354 29  | W44293 | Human vascular endoth   | 1.45e-02 |
| 14        | 46    | 63.0         | 358 22  | W44295 | Murine vascular endoth  | 1.45e-02 |
| 15        | 46    | 63.0         | 358 22  | W14992 | Murine c-Fos induced    | 1.45e-02 |
| 16        | 46    | 63.0         | 592 19  | P96347 | Malic enzyme #2.        | 1.45e-02 |
| 17        | 46    | 63.0         | 630 22  | W14594 | Human c-Fos induced g   | 1.45e-02 |
| 18        | 46    | 63.0         | 749 2   | P70286 | Human c-Fos induced g   | 1.45e-02 |

|    |    |      |       |    |         |                        |           |
|----|----|------|-------|----|---------|------------------------|-----------|
| 19 | 46 | 63.0 | 1.330 | 3  | 1R3444  | Swine herpes virus-1   | 1.45e+020 |
| 20 | 45 | 61.6 | 218   | 5  | R22335  | Sequence encoded by p  | 1.88e+020 |
| 21 | 45 | 61.6 | 251   | 28 | W37358  | Hsf protein involved   | 1.88e+020 |
| 22 | 45 | 61.6 | 476   | 19 | W00630  | ILTV protein kinase    | 1.88e+020 |
| 23 | 45 | 61.6 | 736   | 26 | W19000  | Duck parvovirus capsid | 1.88e+020 |
| 24 | 45 | 61.6 | 1537  | 27 | W34624  | Human C3 protein mut   | 1.88e+020 |
| 25 | 45 | 61.6 | 1635  | 27 | W34629  | Human C3 protein mut   | 1.88e+020 |
| 26 | 45 | 61.6 | 1661  | 27 | W34625  | Human C3 protein mut   | 1.88e+020 |
| 27 | 45 | 61.6 | 1663  | 27 | W34609  | Human C3 protein mut   | 1.88e+020 |
| 28 | 45 | 61.6 | 1663  | 27 | W34610  | Human C3 protein mut   | 1.88e+020 |
| 29 | 45 | 61.6 | 1663  | 27 | W34608  | Human C3 protein mut   | 1.88e+020 |
| 30 | 45 | 61.6 | 1663  | 27 | W34607  | Human C3 protein mut   | 1.88e+020 |
| 31 | 45 | 61.6 | 1663  | 27 | W34606  | Wild type human C3 pr  | 1.88e+020 |
| 32 | 45 | 61.6 | 1663  | 27 | W346090 | Human C3 protein mut   | 1.88e+020 |
| 33 | 45 | 61.6 | 1663  | 27 | W34628  | Human C3 protein mut   | 1.88e+020 |
| 34 | 45 | 61.6 | 1663  | 27 | W34627  | Human C3 protein mut   | 1.88e+020 |
| 35 | 45 | 61.6 | 1663  | 27 | W340989 | Human C3 protein mut   | 1.88e+020 |
| 36 | 45 | 61.6 | 1663  | 27 | W34612  | Human C3 protein mut   | 1.88e+020 |
| 37 | 45 | 61.6 | 1663  | 27 | W34611  | Human C3 protein mut   | 1.88e+020 |
| 38 | 45 | 61.6 | 1663  | 27 | W34620  | Human C3 protein mut   | 1.88e+020 |
| 39 | 45 | 61.6 | 1663  | 27 | W34620  | Human C3 protein mut   | 1.88e+020 |
| 40 | 45 | 61.6 | 1663  | 27 | W34618  | Human C3 protein mut   | 1.88e+020 |
| 41 | 45 | 61.6 | 1663  | 27 | W34617  | Human C3 protein mut   | 1.88e+020 |
| 42 | 45 | 61.6 | 1663  | 27 | W34613  | Human C3 protein mut   | 1.88e+020 |
| 43 | 45 | 61.6 | 1663  | 27 | W34614  | Human C3 protein mut   | 1.88e+020 |
| 44 | 45 | 61.6 | 1667  | 27 | W34631  | Human C3 protein mut   | 1.88e+020 |
| 45 | 45 | 61.6 | 1667  | 27 | W34626  | Human C3 protein mut   | 1.88e+020 |

## ALIGNMENTS

|        |                                |
|--------|--------------------------------|
| RESULT | 1                              |
| ID     | R89369 standard; peptide; 9 AA |

DT 18-SEP-1996 (first entry

DE Cw6 consensus peptide derived immunogenic peptide #1  
KW Immunogenic peptide; supermotif; HLA molecule; CTL

KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.

OS Synthetic.  
PN WO9603140-A1

PD 08-FEB-1996.  
PF 21-JUL-1995; U09234

PR 21-JUL-1994; US-278634  
PR 23-NOV-1994; US-344824

PR 30-MAY-1995; US-452843  
PA (CYTE-) CYTEL CORP.

PI Sette A, Sidney J,  
DR WPI; 96-116784/12.

PT Compsn. comprising  
PT than one HLA mol. t

PT and for in vivo and  
PS Claim 2; page 26; 3

cc The sequences given  
cc use in the composition

allows binding of

```
CC two conserved residues
CC terminal 1s Pro, and
CC
```

CC are used to induce  
CC useful in compositi

CC diagnostic applications,  
CC infections, e.g. hepatitis

sequence 9 AA;

Query Match  
Best Local Similarity

Matches y; Conser

DB 1 1pyplvrkl 9  
| | | | |  
| | | | |

1 IPYPIVKKL 9

RESULT 2  
ID R89370 standard; peptide; 9 AA.  
AC R89370.  
DT 18-SEP-1996 (first entry)  
DE Cw6 consensus peptide derived immunogenic peptide #2.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI: 96-116784/12.  
PT Compsn. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were use in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which allows binding of more than one HLA molecule. It pref. comprises two conserved residues, a first at the 2nd position from the N-terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also useful in compositions for in vivo and ex vivo therapeutic and diagnostic applications, e.g. the treatment of cancer and viral infections, e.g. hepatitis B and C.  
CC Sequence 9 AA;

Query Match 89.0%; Score 65; DB 18; Length 9;  
Best Local Similarity 88.9%; Pred. No. 7.54e-01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 lpylvrsl 9  
1 lpylvrsl 9  
1 lpylvrsl 9

RESULT 3  
ID R89371 standard; peptide; 9 AA.  
AC R89371;  
DT 18-SEP-1996 (first entry)  
DE Cw6 consensus peptide derived immunogenic peptide #3.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI: 96-116784/12.  
PT Compsn. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were use in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which allows binding of more than one HLA molecule. It pref. comprises two conserved residues, a first at the 2nd position from the N-terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also

CC useful in compositions for in vivo and ex vivo therapeutic and CC diagnostic applications, e.g. the treatment of cancer and viral CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 72.6%; Score 53; DB 18; Length 9;  
Best Local Similarity 85.7%; Pred. No. 2.23e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 lpylvr 7  
1 lpylvr 7  
1 lpylvr 7

RESULT 4  
ID W28511 standard; Protein; 133 AA.  
AC W28511;  
DT 29-DEC-1997 (first entry)  
DE Product of clone L105.  
KW J5; J422; L105; H174-10; H174-43; B18; cytokine; PMBC;  
KW peripheral blood mononuclear cell; disintegrin; metallo-protein;  
KW Desophila; leucine-rich repeat; monocyte; chemoattractant;  
KW IF-10; CRG-2; CTLA-8; herpesvirus; Salmi.  
OS Mus musculus.  
PN WO9707198-A2.  
PD 27-FEB-1997.  
PF 08-AUG-1996; U12897.  
PR 08-AUG-1996; WO-U12897.  
PA (GEMV) GENETICS INST INC.  
PI Carlin M, Jacobs K, Kelleher K, McCoy JM;  
DR WPI: 97-165283/15.  
DR N-PSDB; T87429.  
PT Polynucleotide(s) encoding proteins for treating, preventing and ameliorating medical conditions - obtained from human activated peripheral blood mononuclear cell, and murine adult thymus libraries  
PS Claim 21; Page 44-45; 61pp; English.  
CC This sequence was isolated from a murine adult thymus library using a trap selecting for nucleotides encoding secreted proteins, and encodes a protein having homology to various monocyte and other chemottractant proteins.  
CC Sequence 133 AA;

Query Match 69.9%; Score 51; DB 25; Length 133;  
Best Local Similarity 85.7%; Pred. No. 3.85e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 40 lpylvr 46  
1 lpylvr 7  
1 lpylvr 7

RESULT 5  
ID P70388 standard; protein; 356 AA.  
AC P70388;  
DT 14-JAN-1991 (first entry)  
DE D-amino acid oxidase.  
KW D-amino acid oxidase; Trigonopsis variabilis; cephalosporin;  
KW oxidative deamination.  
OS Trigonopsis variabilis.  
PN J62262994-A.  
PD 16-NOV-1987.  
PF 12-MAY-1986; J06663.  
PR 12-MAY-1986; JP-106663.  
PA (ASAHI) ASAHI CHEMICAL IND KK.  
DR WPI: 87-359677/51.  
DR N-PSDB; N70609.  
PT DNA fragment encoding D-amino acid oxidase - which is a useful enzyme for the catalytic oxidative deamination of D-amino acids.  
PS Claim 1; Page 583-4; 12pp; Japanese.  
CC D-amino acid oxidase catalyses the oxidative deamination of D-amino acids. It is used in the sepn. of L-amino acids from racemates, in the prepn. of ketoic acid from D-amino acid, in amino acid analysis, etc. The enzyme can oxidise cephalosporin C to

CC 7-beta-(5-carboxy-5-oxopentanamide)cephalosporanic acid, which  
 CC reacts with hydrogen peroxide to give 7-beta-(4-carboxybutanamide)-  
 CC cephalosporanic acid. These cpds. are important intermediates for  
 CC synthesis of cephalosporin type antibiotics.  
 SQ Sequence 356 AA;

Query Match 68.5%; Score 50; DB 2; Length 356;  
 Best Local Similarity 55.6%; Pred. No. 5.04e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 66 vsypilrel 74  
 :|||:|  
 QY 1 IPPYIVRKL 9

RESULT 6  
 ID R04066 standard: Protein; 356 AA.  
 AC R04066;  
 DT 03-SEP-1990 (first entry)  
 DE T-variabilis D-amino acid oxidase gene product.  
 CC D-amino acid oxidase; cephalosporin; cephem; E.coli.  
 OS Trigonopsis variabilis.  
 PN RP-364275-A.  
 PD 18-APR-1990.  
 PF 12-OCT-1989; 310483.  
 PR 13-OCT-1988; JP-260332.  
 PA (FUJII) Fujisawa Pharm KK.  
 PI Isoval T, Ono H, Kojo H;  
 DR WPI; 90-117771/16.  
 PT D-amino acid oxidase, prodn.  
 PT by culture of E.coli transformants contg. expression vectors  
 PT originated from Fusarium solani M-0718.  
 PS Disclosure: Fig 9; 38pp; English.  
 CC E.coli transformed to express DAO, which catalyses the enzymatic  
 CC conversion of cephalosporin C to 7-beta-(5-carboxy-5-  
 CC oxopentanamido)cephalosporanic acid (keto-7ACA). 7ACA is an  
 CC important starting point for the production of cephem  
 CC antibiotics.  
 SQ Sequence 356 AA;

Query Match 68.5%; Score 50; DB 3; Length 356;  
 Best Local Similarity 55.6%; Pred. No. 5.04e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 65 vsypilrel 73  
 :|||:|  
 QY 1 IPPYIVRKL 9

RESULT 7  
 ID R77604 standard: Protein; 1676 AA.  
 AC R77604;  
 DT 15-MAR-1996 (first entry)  
 DE Pro-C5 polypeptide.  
 CC Complement C5; haemolysis; kidney; glomerulonephritis;  
 KW monoclonal antibody; antiinflammatory; antibody engineering;  
 KW humanised antibody.  
 OS Homo sapiens.  
 FH Key  
 FT Peptide  
 FT 1..18  
 FT Location/Qualifiers  
 FT /label= Sig\_peptide  
 FT 19..673  
 FT /label= Beta-chain  
 FT 673..674  
 FT cleavage\_site  
 FT 677..678  
 FT cleavage\_site  
 FT 674..677  
 FT label= cleavage\_peptide  
 FT 678..1676  
 FT /label= Alpha-chain  
 FT /note= "Amino acids 872-892 (854-874 of  
 FT the mature protein) comprise the KSSKS  
 FT epitope"  
 FT 678..751  
 FT peptide

FT /label= C5a  
 FT 751..752  
 FT /label= Convertase\_cleavage\_site  
 FT modified\_site  
 FT 911  
 FT /label= N-glycosylation\_site  
 FT modified\_site  
 FT 1115  
 FT /label= N-glycosylation\_site  
 FT modified\_site  
 FT 1630  
 FT /label= N-glycosylation\_site

PN W09529697-A1.  
 PD 09-NOV-1995.  
 PE 01-MAY-1995; U05688.  
 PR 02-MAY-1994; US-236208.  
 PA (ALEX-) ALEXION PHARM INC.  
 PI Evans MJ, Matis L, Mueller EE, Nye SH, Rollins S;  
 PI Rotner RP, Springhorn J P, Sguinto SP, Thomas TC;  
 PI Wang Y, Wilkins JA;  
 DR WPI; 95-392923/50.  
 PT Treating glomerulonephritis with antibody against complement C5  
 PT component - to inhibit complement induced cell lysis  
 PS Example 13; Page 82-92; 181pp; English.  
 CC The cDNA sequence of the complement C5 gene transcript predicts a  
 CC secreted pro-C5 precursor of 1676 amino acids (R77604). C5 is a  
 CC beta-globulin heterodimer thought to play a role in the pathogenesis  
 CC of glomerulonephritis (GN). Cleavage of the C5 alpha-chain  
 CC by a convertase enzyme generates anaphylatoxic C5a. Monoclonal  
 CC and humanised recombinant antibodies that recognise the alpha-chain  
 CC KSSKC epitope (R77605) block C5a generation, thereby reducing  
 CC glomerular inflammation and kidney dysfunction associated with GN.  
 SQ Sequence 1676 AA;

Query Match 67.1%; Score 49; DB 15; Length 1676;  
 Best Local Similarity 71.4%; Pred. No. 6.58e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 829 lpsvvr 835  
 |||||:  
 QY 1 IPPYIVR 7

RESULT 8  
 ID W44125 standard: peptide; 159 AA.  
 AC W44125;  
 DT 28-APR-1998 (first entry)  
 DE Streptococcus pneumoniae leucyl tRNA synthetase fragment.  
 KW leucyl tRNA synthetase; leucyl polypeptide; vaccine;  
 KW genetic immunisation; antibacterial; antibiotic; otitis media;  
 KW conjunctivitis; pneumonia; bacteremia; meningitis; sinusitis;  
 KW pleural emphysema; endocarditis; gene therapy.  
 OS Streptococcus pneumoniae.  
 FH Key  
 FT Location/Qualifiers  
 FT Misc\_difference 53  
 FT /label= unspecified  
 FT /note= "encoded by NNA"  
 PN W09739022-A1.  
 PD 23-OCT-1997.  
 PE 18-APR-1997; U06875.  
 PR 18-APR-1996; GB-007993.  
 PA (SMK ) SMITHKLINE BEECHAM CORP.  
 PA (SMK ) SMITHKLINE BEECHAM PLC.  
 PI Laviot EJ;  
 PI WPI; 97-526396/48.  
 DR N-PSDB; V12059.  
 PT Streptococcus pneumoniae leucyl tRNA synthetase - useful to produce  
 PT antibodies or to screen for (ant)agonists with antibacterial  
 PT activity, e.g. to diagnose and treat meningitis, pneumonia, etc.  
 PS Claim 12; Page 39-40; 48pp; English.  
 CC The present sequence represents a leucyl tRNA synthetase (leus) fragment  
 CC from Streptococcus pneumoniae. The leus polypeptides, antagonists,  
 CC antibodies and related nucleic acids can be used for diagnosis and  
 CC treatment of bacterial diseases. In particular, they are directed  
 CC towards Streptococcus pneumoniae infections causing otitis media,  
 CC conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis, pleural

region

FT /label= Helix-C  
 FT /note= "amphipathic helix has a repeat pattern of  
 FT apolar residues in the 1 and 1+3 positions  
 FT of a heptad  
 FT  
 FT disulfide\_bond  
 FT 135  
 FT /note= "cysteine residue involved in disulphide  
 FT bonding"  
 FT 169..191  
 FT region  
 FT /label= Helix-D  
 FT /note= "amphipathic helix has a repeat pattern of  
 FT apolar residues in the 1 and 1+3 positions  
 FT of a heptad"  
 FT 169..176  
 FT region  
 FT /label= D1  
 FT /note= "conserved D1 region is implicated in  
 FT receptor binding"  
 FT  
 FT disulfide\_bond  
 FT 177  
 FT /note= "cysteine residue involved in disulphide  
 FT bonding"  
 FT 185..188  
 FT region  
 FT /label= D2  
 FT /note= "D2 region implicated in receptor binding"  
 FT 207..245  
 FT /label= COOH\_extension  
 FT  
 FT region  
 FT  
 FT EP-748870-AZ.  
 FT 18-DEC-1996.  
 FT 05-JUN-1996; 109045.  
 FT 06-JUN-1995; US-473026.  
 FT (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 FT C1e99 CH;  
 FT MPI: 97-036153/04.  
 FT N-PSDB: T43276.  
 FT  
 FT Haematopoietic stem cell proliferation and differentiation using  
 FT oncostatin M - useful for increasing survival of lymphocytes during  
 FT treatment of e.g. lymphoma, leukaemia or AIDS  
 FT Example 1; Fig 1B; 18pp; English.  
 FT  
 FT Bovine oncostatin M (M08080) was identified as the product of a  
 FT gene (T43276) isolated from a bovine genomic library using human  
 FT oncostatin M cDNA as a probe. Expression of oncostatin M in  
 FT transfectant haematopoietic stem cells can lead to an induction of  
 FT proliferation and differentiation of the stem cells. This can be  
 FT utilised in the treatment of e.g. lymphoma, leukaemia and AIDS.  
 FT Expression of oncostatin M in transfectant stem cells may also be  
 FT used to induce haematopoiesis in a mammal, thereby prolonging  
 FT lymphocyte survival upon exposure to cytotoxic agents such as  
 FT mitomycin C or radiation.  
 FT  
 FT Sequence 245 AA;  
 SQ

Query Match 63.0%; Score 46; DB 20; Length 245;  
 Best Local Similarity 50.0%; Pred. No. 1.45e+02;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 222 pfelr1 229  
 1:1:1:1  
 Qy 2 PYPIVRKL 9

RESULT 12  
 ID W44296 standard; Protein; 326 AA.  
 AC W44296;  
 DE 22-JUN-1998 (first entry)  
 DE Rat vascular endothelial growth factor D.  
 KW Rate; vascular endothelial growth factor D; VEGF-D; gene therapy;  
 KW Inflammation; oedema.  
 OS Rattus sp.  
 PN M09802543-A1.  
 PD 22-JAN-1998.  
 PF 15-JUL-1997; J02456.  
 PR 15-JUL-1996; JP-185216.  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 PI Hirata Y, Nezu J;  
 DR MPI: 98-110591/10.  
 DR N-PSDB: V15176.

PT VEGF-D protein encoded by DNA - useful for, e.g. gene therapy and  
 PT treating oedema  
 PS Example 8; Page 35-38; 52pp; Japanese.  
 CC The present sequence represents rat vascular endothelial growth factor D  
 CC (VEGF-D). The VEGF-D protein, compounds and antibodies, which can bind  
 CC the protein, may be useful in, e.g. gene therapy and in treatment of  
 CC inflammation and oedema. Vectors, containing the VEGF-D DNA, and VEGF-D  
 CC DNA sequences may be used for screening for the compounds which bind to  
 CC the VEGF-D protein.  
 SQ Sequence 326 AA;

Query Match 63.0%; Score 46; DB 29; Length 326;  
 Best Local Similarity 57.1%; Pred. No. 1.45e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 204 pyp1vrk 210  
 1:1:1:1  
 Qy 2 PYPIVRK 8

RESULT 13  
 ID W44293 standard; Protein; 354 AA.  
 AC W44293;  
 DE 22-JUN-1998 (first entry)  
 DE Human vascular endothelial growth factor D.  
 KW Human; vascular endothelial growth factor D; VEGF-D; gene therapy;  
 KW Inflammation; oedema.  
 OS Homo sapiens.  
 PN M09802543-A1.  
 PD 22-JAN-1998.  
 PF 15-JUL-1997; J02456.  
 PR 15-JUL-1996; JP-185216.  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 PI Hirata Y, Nezu J;  
 DR MPI: 98-110591/10.  
 DR N-PSDB: V15156.  
 PT VEGF-D protein encoded by DNA - useful for, e.g. gene therapy and  
 PT treating oedema  
 PS Claim 1; Page 18-20; 52pp; Japanese.  
 CC The present sequence represents human vascular endothelial growth factor  
 CC D (VEGF-D). The VEGF-D protein, compounds and antibodies, which can bind  
 CC the protein, may be useful in, e.g. gene therapy and in treatment of  
 CC inflammation and oedema. Vectors, containing the VEGF-D DNA, and VEGF-D  
 CC DNA sequences may be used for screening for the compounds which bind to  
 CC the VEGF-D protein.  
 SQ Sequence 354 AA;

Query Match 63.0%; Score 46; DB 29; Length 354;  
 Best Local Similarity 57.1%; Pred. No. 1.45e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 199 pyp1vrk 205  
 1:1:1:1  
 Qy 2 PYPIVRK 8

RESULT 14  
 ID W44295 standard; Protein; 358 AA.  
 AC W44295;  
 DE 22-JUN-1998 (first entry)  
 DE Mouse vascular endothelial growth factor D.  
 KW Mouse; vascular endothelial growth factor D; VEGF-D; gene therapy;  
 KW Inflammation; oedema.  
 OS Mus sp.  
 PN M09802543-A1.  
 PD 22-JAN-1998.  
 PF 15-JUL-1997; J02456.  
 PR 15-JUL-1996; JP-185216.  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 PI Hirata Y, Nezu J;  
 DR MPI: 98-110591/10.  
 DR N-PSDB: V15177.  
 PT VEGF-D protein encoded by DNA - useful for, e.g. gene therapy and

PT Treating oedema  
 PS Example 7; Page 32-35; 52pp; Japanese.  
 CC The present sequence represents mouse vascular endothelial growth factor  
 CC D (VEGF-D). The VEGF-D protein, compounds and antibodies, which can bind  
 CC the protein, may be useful in, e.g. gene therapy and in treatment of  
 CC inflammation and oedema. Vectors, containing the VEGF-D DNA, and VEGF-D  
 CC DNA sequences may be used for screening for the compounds which bind to  
 CC the VEGF-D protein.  
 SQ Sequence 358 AA;

Query Match: 63.0%; Score 46; DB 29; Length 358;  
 Best Local Similarity 57.1%; Pred. No. 1.45e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

DB 204 pvsllrr 210  
 11:11:  
 OY 2 pypivrk 8

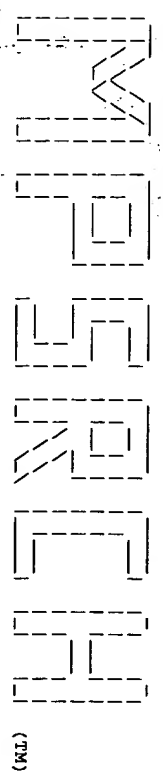
## RESULT 15

ID W14992 standard; Protein; 358 AA.  
 AC W14992;  
 DT 05-JUL-1997 (first entry)  
 DE Murine c-Fos induced growth factor.  
 KW c-Fos induced growth factor; FIGF; Fos regulated gene;  
 KW proto-oncogene; lung disorder; cancer; tumour; therapy;  
 KW antibody; transgenic animal.  
 OS Mus sp.  
 FH Key  
 FT Location/Qualifiers  
 FT region 112..164  
 FT /note= "VEGF homology region"  
 PN WO9712972-A2.  
 PD 10-APR-1997.  
 PF 30-SEP-1996; IB1113.  
 PR 29-SEP-1995; GB-01928.  
 PR 13-JUN-1996; GB-012368.  
 PA 1 (UYSI-) UNIV SIENA.  
 PI Oliviero S;  
 DR MPI: 97-226216/20.  
 DR N-PSDB; T62960.  
 PT Nucleotide molecule encoding c-Fos induced growth factor protein -  
 PT useful in therapy, in manufacture of compositions for treatment of  
 PT developmental disorders and in generation of transgenic animal  
 PS Claim 3; Fig 1; 64pp; English.  
 CC Novel murine c-Fos-induced growth factor (FIGF) (W14992) shows  
 CC homology to the growth factor VEGF. It is encoded by the F0401  
 CC gene (T62960) obtd. from mouse fibroblast cells. FIGF is a c-fos-  
 CC dependent autocrine growth factor able to induce cell division  
 CC entry and, when over-expressed, a transformed phenotype in  
 CC fibroblasts. It could be implicated in tumours and development.  
 CC Recombinant FIGF can be produced in transformed host (e.g. CHO)  
 CC cells. It can be used to identify its receptors and in an assay  
 CC for the identification of agonists and antagonists. Antibodies  
 CC raised against FIGF can be used to block the function of the  
 CC protein and thereby inhibit or suppress tumour growth. Transgenic  
 CC animals expressing FIGF can be generated for use e.g. as models for  
 CC research.  
 SQ Sequence 358 AA;

Query Match 63.0%; Score 46; DB 22; Length 358;  
 Best Local Similarity 57.1%; Pred. No. 1.45e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 DB 204 pvsllrr 210  
 11:11:  
 OY 2 pypivrk 8

Search completed: Fri Sep 11 13:06:00 1998  
 Job time : 16 secs.





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_p protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Sep 11 13:06:17 1998; MasPar time 3.04 Seconds  
Tabular output not generated. 108.312 Million cell updates/sec

Title: >US-08-452-843-8  
Description: (1-9) from US08452843.pep  
Perfect Score: 73  
Sequence: 1 IPYPIVKRL 9

Scoring table:  
PAM 150  
Gap 15

Searched: 120441 seqs, 36531193 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r56  
1: p1r1 2: p1r2 3: p1r3 4: p1r4 5: r1r13d

Statistics: Mean 23.889; Variance 33.119; scale 0.721

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description           | Pred. No. |
|------------|-------|-------------|--------|----|--------|-----------------------|-----------|
| 1          | 55    | 75.3        | 1679   | 2  | S48385 | hypothetical protein  | 2.58e+00  |
| 2          | 53    | 72.6        | 52     | 2  | F70083 | hypothetical protein  | 6.12e+00  |
| 3          | 53    | 72.6        | 209    | 2  | S13179 | transforming protein  | 6.12e+00  |
| 4          | 53    | 72.6        | 487    | 1  | OQECRS | hypothetical yvjf pro | 6.12e+00  |
| 5          | 52    | 71.2        | 451    | 2  | S75569 | hypothetical protein  | 9.36e+00  |
| 6          | 52    | 71.2        | 880    | 2  | B33926 | DNA-directed RNA poly | 9.36e+00  |
| 7          | 51    | 69.9        | 75     | 2  | H69915 | hypothetical protein  | 1.42e+01  |
| 8          | 51    | 69.9        | 197    | 2  | C64422 | hypothetical protein  | 1.42e+01  |
| 9          | 51    | 69.9        | 340    | 2  | S62493 | hypothetical protein  | 1.42e+01  |
| 10         | 51    | 69.9        | 607    | 1  | NUTR8  | glucose-6-phosphate 1 | 1.42e+01  |
| 11         | 51    | 69.9        | 766    | 2  | A56394 | pyocin S3 Pseudomon   | 1.42e+01  |
| 12         | 50    | 68.5        | 615    | 2  | H64769 | secretion protein sec | 2.15e+01  |
| 13         | 49    | 67.1        | 125    | 2  | S23541 | hypothetical protein  | 3.24e+01  |
| 14         | 49    | 67.1        | 183    | 2  | B64506 | adenine phosphoribosy | 3.24e+01  |
| 15         | 49    | 67.1        | 227    | 2  | C69432 | hypothetical protein  | 3.24e+01  |
| 16         | 49    | 67.1        | 1061   | 1  | S27311 | ribonuclease E (EC 3  | 3.24e+01  |
| 17         | 49    | 67.1        | 1676   | 1  | C5HU   | complement C5 precurs | 3.24e+01  |
| 18         | 49    | 67.1        | 1680   | 1  | C5MS   | complement C5 precurs | 3.24e+01  |
| 19         | 48    | 65.8        | 100    | 2  | S44892 | ZK112.4 protein - Cae | 4.84e+01  |
| 20         | 48    | 65.8        | 156    | 2  | C64486 | hypothetical protein  | 4.84e+01  |
| 21         | 48    | 65.8        | 219    | 2  | S75541 | hypothetical protein  | 4.84e+01  |
| 22         | 48    | 65.8        | 303    | 2  | JN0857 | C-alpha-dehydrogenase | 4.84e+01  |
| 23         | 48    | 65.8        | 305    | 2  | S35991 | C-alpha dehydrogenase | 4.84e+01  |

|    |    |      |      |   |        |                        |          |
|----|----|------|------|---|--------|------------------------|----------|
| 24 | 48 | 65.8 | 315  | 2 | D64127 | htrb protein - Haemop  | 4.84e+01 |
| 25 | 48 | 65.8 | 326  | 2 | S26216 | glutamate--ammonia li  | 4.84e+01 |
| 26 | 48 | 65.8 | 377  | 2 | S25156 | transposase - Bacilli  | 4.84e+01 |
| 27 | 48 | 65.8 | 407  | 2 | B32306 | cytochrome P450 104 -  | 4.84e+01 |
| 28 | 48 | 65.8 | 540  | 2 | A55145 | thiamin-phosphate pyr  | 4.84e+01 |
| 29 | 48 | 65.8 | 617  | 2 | A39748 | nerve growth factor-1  | 4.84e+01 |
| 30 | 48 | 65.8 | 617  | 2 | I56530 | gene VGF protein - ra  | 4.84e+01 |
| 31 | 48 | 65.8 | 762  | 1 | NNNC2  | anthranilate synthase  | 4.84e+01 |
| 32 | 48 | 65.8 | 1381 | 2 | S55619 | capsid protein 25 - e  | 4.84e+01 |
| 33 | 47 | 64.4 | 196  | 2 | B69042 | conserved hypothetical | 7.19e+01 |
| 34 | 47 | 64.4 | 238  | 2 | B64313 | hypothetical protein   | 7.19e+01 |
| 35 | 47 | 64.4 | 268  | 2 | S39711 | conserved hypothetical | 7.19e+01 |
| 36 | 47 | 64.4 | 329  | 1 | AJZ202 | glutamate--ammonia li  | 7.19e+01 |
| 37 | 47 | 64.4 | 329  | 1 | WAFB4  | 1329L protein - Altic  | 7.19e+01 |
| 38 | 47 | 64.4 | 364  | 2 | S77360 | hypothetical protein   | 7.19e+01 |
| 39 | 47 | 64.4 | 381  | 2 | C64416 | hypothetical protein   | 7.19e+01 |
| 40 | 47 | 64.4 | 472  | 2 | S62494 | hypothetical protein   | 7.19e+01 |
| 41 | 47 | 64.4 | 682  | 1 | HHBYK2 | dnak-type molecular c  | 7.19e+01 |
| 42 | 47 | 64.4 | 682  | 1 | S76532 | hypothetical protein   | 7.19e+01 |
| 43 | 47 | 64.4 | 1036 | 2 | S55151 | probable membrane pro  | 7.19e+01 |
| 44 | 47 | 64.4 | 1868 | 2 | S48938 | hypothetical protein   | 7.19e+01 |
| 45 | 46 | 63.0 | 380  | 5 | ICXSA1 | dimeethylsulfoxide red | 1.06e+02 |

ALIGNMENTS

RESULT 1  
ENTRY S48385 #type complete  
TITLE hypothetical protein YJL149c - yeast (Saccharomyces cerevisiae)  
ORGANISM #formal\_name Saccharomyces cerevisiae  
DATE 02-Dec-1994 #sequence\_revision 02-Dec-1994 #text\_change 21-Nov-1997  
ACCESSIONS S48385  
REFERENCE S48310  
#authors Churcher, C.  
#submission submitted to the EMBL Data Library, September 1994  
#accession S48385  
#molecule\_type DNA  
#residues 1-1679 #label CHU  
#cross-references GB:247047; EMBL:238059; NID:g603997; PID:g763197; MIPS:YJL149c  
GENETICS  
#map\_position 9L  
SUMMARY #length 1679 #molecular-weight 195140 #checksum 5165  
Query Match 75.3%; Score 55; DB 2; Length 1679;  
Best Local Similarity 55.6%; Pred. No. 2.58e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
DB 19 VTPYVLRKL 27  
QY : ||: |||  
1 IPYPIVKRL 9  
RESULT 2  
ENTRY F70083 #type complete  
TITLE hypothetical protein yxzf - Bacillus subtilis  
ORGANISM #formal\_name Bacillus subtilis  
DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Dec-1997  
ACCESSIONS F70083  
REFERENCE A69580  
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allion, G.; Azevedo, V.; Berto, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denicot, F.; Devine, K.M.; Duesterhoft, A.; Ehrlich, S.D.; Emmerson, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,

S. Y.; Glaser, P.; Goffeau, A.; Golightly, E. J.; Grandi, G.; Guisepi, G.; Guy, B. J.; Haga, K.; Heideh, J.; Harwood, C. R.; Hentut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M. F.; Itaya, M.; Jones, L.; Joris, B.; Katamata, D.; Kasahara, Y.; Klier-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koelter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino, S.; Lauber, J.; Lazarevic, V.; Lee, S. M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R. P.; Mizuno, M.; Moestl, D.; Nakai, S.; Nodack, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudea, B.; Park, S. H.; Parro, V.; Pohl, T. M.; Portelle, D.; Porvolik, S.; Prescott, A. M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scorfone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S. J.; Serro, P.; Shin, B. S.; Soldo, B.; Sorokin, A.; Taccini, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wanduit, R.; Wedler, E.; Wedler, H.; Wetzemeyer, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H. F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

#journal  
#title Nature (1997) 390:249-256  
#accession F70083  
#status preliminary; nucleic acid sequence not shown;  
translation not shown

##molecule-type DNA  
##residues 1-52 #label KUN  
##experimental-source strain 168

GENETICS  
#gene yzf  
SUMMARY  
#length 52 #molecular-weight 5915 #checksum 5675

Query Match  
Best Local Similarity 66.7%; Score 53; DB 2; Length 52;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 25 IIPYIVRL 33  
: |||||:  
OY 1 IIPYIVRL 9

RESULT 3  
ENTRY S13179 #type complete  
TITLE transforming protein (ras) - Geodia cydonium  
ORGANISM #formal\_name Geodia cydonium  
DATE 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 31-Oct-1997

ACCESSIONS  
REFERENCE S13179  
#authors Robitzki, A.; Schroeder, H. C.; Ugarkovic, D.; Kuchino, Y.; Kurelec, B.; Gamulin, V.; Mueller, W. E. G.  
#journal Eur. J. Biochem. (1990) 192:499-506  
#title Regulated expression and phosphorylation of the 23-26-kDa ras protein in the sponge Geodia cydonium.  
#cross-references MUID:91006138  
#accession S13179  
##status preliminary  
##molecule-type mRNA  
##residues 1-209 #label ROB  
#note based on the evidence for Gln-tRNA, the authors translated the codon TAG as Gln; the sequence shown follows the authors' translation

CLASSIFICATION  
SUMMARY #superfamily ras transforming protein  
#length 209 #molecular-weight 23854 #checksum 3860

Query Match  
Best Local Similarity 66.7%; Score 53; DB 2; Length 209;  
Matches 6; Conservative 6; Pred. No. 6.12e+00;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 178 IIPYIVRL 186  
: |||||:  
OY 1 IIPYIVRL 9

RESULT 4  
ENTRY Q06CRS #type complete  
TITLE hypothetical yjye protein - Escherichia coli (strain K-12)  
ORGANISM #formal\_name Escherichia coli  
DATE 30-Jun-1998 #sequence\_revision 31-Oct-1997 #text\_change 14-Nov-1997

ACCESSIONS  
REFERENCE E65094; C29049  
#authors A64720  
Blattner, F. R.; Plunkett III, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. R.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y.

#journal  
#title Science (1997) 277:1453-1462  
#cross-references MUID:97426617  
#accession E65094  
#status nucleic acid sequence not shown; translation not shown

##molecule-type DNA  
##residues 1-487 #label BIAT  
##cross-references GB:AEO00388; GB:U00096; NID:91789441; PID:91789444; UMG:P:D3063

REFERENCE  
#experimental-source strain K-12, substrain MG1655  
A91573  
Nesin, M.; Lupski, J. R.; Svec, P.; Godson, G. N.  
#journal Gene (1987) 51:149-161  
#title Possible new genes as revealed by molecular analysis of a 5-kb Escherichia coli chromosomal region 5' to the rpsu-dnaG-rpoD macromolecular-synthesis operon.  
#cross-references MUID:87248073  
#accession C29049

##molecule-type DNA  
##residues 279-403, 'P', 405-411, 'RWCRKRSRCRCSA' #label NES

GENETICS  
#gene yjye  
#map\_position 67 min  
SUMMARY  
#length 487 #molecular-weight 52906 #checksum 1643

Query Match  
Best Local Similarity 66.7%; Score 53; DB 1; Length 487;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 165 IIPYIVRL 173  
: |||||:  
OY 1 IIPYIVRL 9

RESULT 5  
ENTRY S75569 #type complete  
TITLE hypothetical protein slr0818 - Synecocystis sp. (PCC 6803)  
ORGANISM #formal\_name Synecocystis sp.  
#variety PCC 6803  
DATE 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Sep-1997

ACCESSIONS  
REFERENCE S75569  
#authors S74322  
Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugitara, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Nartuo, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tadeta, S.  
#journal DNA Res. (1996) 3:109-136  
#title Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of

Potential protein-coding regions.

#cross-references MUID:97061201  
#accession S75569

##status Preliminary  
##molecule\_type DNA  
##residues 1-451 #label KAN  
#cross-references EMBL:D90911; NID:g1653083; PID:d1018863; PID:g1653214  
#note the nucleotide sequence was submitted to the EMBL Data Library June 1996

SUMMARY #length 451 #molecular-weight 50417 #checksum 4508

Query Match 71.2%; Score 52; DB 2; Length 451;  
Best Local Similarity 62.5%; Pred. No. 9.36e+00;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 110 LPPVRR 117  
1 IPYIVRK 8

RESULT 6  
ENTRY B33926 #type complete  
TITLE DNA-directed RNA polymerase (EC 2.7.7.6) chain A - Sulfolobus acidocaldarius  
ORGANISM #formal\_name Sulfolobus acidocaldarius  
DATE 09-Mar-1990 #sequence\_revision 09-Mar-1990 #text\_change 12-Sep-1997

ACCESSIONS B33926; S04717  
B33926; S04717

REFERENCE  
#authors Puenhler, G.; Jeffers, H.; Grop, F.; Palm, P.; Klenk, H.P.; Lottspeich, F.; Garrett, R.A.; Zillig, W.  
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:4569-4573  
#title Archaeobacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome.  
#cross-references MUID:89282812  
#accession B33926  
#status preliminary; nucleic acid sequence not shown; not compared with conceptual translation

##molecule\_type DNA  
##residues 1-880 #label PUE  
#accession S04714  
#authors Puenhler, G.; Lottspeich, F.; Zillig, W.  
#journal Nucleic Acids Res. (1989) 17:4517-4534  
#title Organization and nucleotide sequence of the genes encoding the large subunits A, B and C of the DNA-dependent RNA polymerase of the archaeobacterium Sulfolobus acidocaldarius.  
#cross-references MUID:89315197  
#accession S04717  
##molecule\_type DNA  
##residues 1-311, 'N', 313-560, 'N', 562-610, 'W', 612-640, 'W', 642-880  
##label P02  
##cross-references EMBL:X14818; NID:g46667; PID:g46670

GENETICS  
#gene rpoA  
#superfamily Halobacterium DNA-directed RNA polymerase chain A

KEYWORDS nucleotidyltransferase; transcription

SUMMARY #length 880 #molecular-weight 99825 #checksum 9710

Query Match 71.2%; Score 52; DB 2; Length 880;  
Best Local Similarity 66.7%; Pred. No. 9.36e+00;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 339 IPYIVRK 347  
1 IPYIVRK 9

RESULT 7  
ENTRY H69915 #type complete  
TITLE hypothetical protein yopB - Bacillus subtilis  
ORGANISM #formal\_name Bacillus subtilis

DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Dec-1997

ACCESSIONS H69915  
REFERENCE A69580  
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denicot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Goldlight, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogn, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinios, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maneel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle, D.; Potwolk, S.; Prescott, A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.; Schlich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, S.J.; Serrin, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassartotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitenegger, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256  
#title The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
#accession H69915  
#status preliminary; nucleic acid sequence not shown; translation not shown

##molecule\_type DNA  
##residues 1-75 #label KUN  
##experimental\_source strain 168

GENETICS  
#gene yopB  
#length 75 #molecular-weight 9099 #checksum 9896

SUMMARY

Query Match 69.9%; Score 51; DB 2; Length 75;  
Best Local Similarity 85.7%; Pred. No. 1.42e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 28 YPTVRKL 34  
3 YPTVRKL 9

RESULT 8  
ENTRY C64422 #type complete  
TITLE hypothetical protein MJ0979 - Methanococcus jannaschii  
ORGANISM #formal\_name Methanococcus jannaschii  
DATE 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 10-Oct-1997

ACCESSIONS C64422  
REFERENCE A64300  
#authors Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;

#title Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodde, A.; Scott, J.L.; Geoghagen, N.S.M.; Meldman, J.F.; Fuhrman, J.L.; Nguyen, D.; Uterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.  
#cross-references M01D:96357999  
#accession C64422  
#status preliminary; nucleic acid sequence not shown; translation not shown

GENETICS  
#map\_position REV912311-911718  
SUMMARY #length 197 #molecular-weight 21520 #checksum 5590

Query Match 69.9%; Score 51; DB 2; Length 197;  
Best Local Similarity 55.6%; Pred. No. 1.42e+01;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 189 IAYPIRVK 197  
|:|:|:|:|:  
QY 1 IYPIVIRKL 9

RESULT 9  
ENTRY S62493 #type complete  
TITLE hypothetical protein SPAC3D3.02 - fission yeast  
ORGANISM (Schizosaccharomyces pombe)  
DATE 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 31-Oct-1997  
ACCESSIONS S62493  
REFERENCE S62492  
#authors Niblett, D.; Harris, D.  
#submission submitted to the EMBL Data Library, October 1995  
#accession S62493  
#status preliminary  
#molecule\_type DNA  
#residues 1-340 #label NIB  
#cross-references EMBL:Z64354; NID:q1039338; PID:q1039340

GENETICS  
#map\_position 1R  
SUMMARY #length 340 #molecular-weight 37876 #checksum 2036

Query Match 69.9%; Score 51; DB 2; Length 340;  
Best Local Similarity 55.6%; Pred. No. 1.42e+01;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 249 VYNYIRSL 257  
:|:|:|:|:  
QY 1 IYPIVIRKL 9

RESULT 10  
ENTRY NUTB #type complete  
TITLE glucose-6-phosphate isomerase (EC 5.3.1.9) - Trypanosoma brucei  
ALTERNATE\_NAMES phosphoglucose isomerase; phosphohexose isomerase  
ORGANISM #formal\_name Trypanosoma brucei  
DATE 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 05-Sep-1997  
ACCESSIONS S06113  
REFERENCE S06113  
#authors Marchand, M.; Kooystra, U.; Wierenga, R.K.; Lambell, A.M.; van Beeumen, J.; Opperdoes, F.R.; Michels, P.A.M.; Bur, J. Biochem. (1989) 184:455-464

#title glucosephosphate isomerase from Trypanosoma brucei. Cloning and characterization of the gene and analysis of the enzyme.  
#cross-references M01D:9005496  
#accession S06113  
#molecule\_type DNA  
#residues 1-607 #label MAR  
#cross-references GB:X15540; NID:q10486; PID:q10487  
#note part of this sequence was confirmed by protein sequencing

CLASSIFICATION #superfamily glucose-6-phosphate isomerase  
KEYWORDS gluconeogenesis; glycolysis; homodimer; intramolecular oxidoreductase; isomerase

FEATURE  
571 #active-site Lys #status predicted  
SUMMARY #length 607 #molecular-weight 67517 #checksum 4076

Query Match 69.9%; Score 51; DB 1; Length 607;  
Best Local Similarity 55.6%; Pred. No. 1.42e+01;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 31 IPEVTRRL 39  
|:|:|:|:|:  
QY 1 IYPIVIRKL 9

RESULT 11  
ENTRY A56394 #type complete  
TITLE pyocin S3 - Pseudomonas aeruginosa (strain P12)  
ORGANISM #formal\_name Pseudomonas aeruginosa  
DATE 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 09-Sep-1997  
ACCESSIONS A56394  
REFERENCE A56394  
#authors Dupont, C.; Baysse, C.; Michel-Briand, Y.  
#journal J. Biol. Chem. (1995) 270:8920-8927  
#title Molecular characterization of pyocin S3, a novel S-type pyocin from Pseudomonas aeruginosa.  
#accession A56394  
#status preliminary  
#molecule\_type DNA  
#residues 1-766 #label DUP  
#cross-references GB:X77996; NID:q854362; PID:q854363

GENETICS  
#gene pyocin S3  
KEYWORDS bacteriocin  
FEATURE  
2-766 #product pyocin S3 #status experimental #label MAT  
SUMMARY #length 766 #molecular-weight 81434 #checksum 2106

Query Match 69.9%; Score 51; DB 2; Length 766;  
Best Local Similarity 66.7%; Pred. No. 1.42e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 679 IYGEIRKL 687  
|:|:|:|:|:  
QY 1 IYPIVIRKL 9

RESULT 12  
ENTRY H64769 #type complete  
TITLE secretion protein seed - Escherichia coli  
ALTERNATE\_NAMES protein-export membrane protein seed  
ORGANISM #formal\_name Escherichia coli  
DATE 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 30-Jan-1998  
ACCESSIONS H64769; J00696; S12301  
REFERENCE A64720  
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

#journal Science (1997) 277:1453-1462  
#title The complete genome sequence of Escherichia coli K-12.  
#cross-references M1000147; GB:97426617  
#accession H64769  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
#molecule-type DNA  
#residues 1-615 ##label BLAT  
#cross-references GB:AE000147; GB:U00096; NID:g1786603; PID:g1786609;  
UMCP:00408  
#experimental-source strain K-12, substrain MG1655  
REFERENCE J00693  
#authors Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.  
#journal EMBO J. (1990) 9:3209-3216  
#title The *secD* locus of *E. coli* codes for two membrane proteins  
required for protein export.  
#cross-references M10006014  
#accession J00696  
#molecule-type DNA  
#residues 1-77,'S',79-154,'A',156-615 ##label GAR  
#cross-references GB:X56175; NID:g42929; PID:g581230  
REFERENCE S12298  
#authors Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.  
#journal EMBO J. (1990) 9:4205-4206  
#contents erratum  
#accession S12301  
#molecule-type DNA  
#residues 1-77,'S',79-154,'A',156-615 ##label GA2  
#cross-references EMBL:X56175; NID:g42929; PID:g581230  
REFERENCE A36969  
#authors Pogliano, K.J.; Beckwith, J.  
#journal J. Bacteriol. (1994) 176:804-814  
#title Genetic and molecular characterization of the *Escherichia coli* *secD* operon and its products.  
#contents annotation; membrane topology  
GENETICS  
#gene *secD*  
#start-codon GTG  
KEYWORDS inner membrane; protein export; transmembrane protein  
FEATURES  
10-30 #domain transmembrane #status predicted #label TM1\  
31-455 #domain periplasmic #status predicted #label P1\  
456-472 #domain transmembrane #status predicted #label TM2\  
477-497 #domain transmembrane #status predicted #label TM3\  
498-501 #domain periplasmic #status predicted #label P2\  
502-518 #domain transmembrane #status predicted #label TM4\  
519-585 #domain periplasmic #status predicted #label P3\  
586-605 #domain transmembrane #status predicted #label TM5\  
#domain transmembrane #status predicted #label TM6  
#length 615 #molecular-weight 66631 #checksum 9609  
SUMMARY  
Query Match 68.5%; Score 50; DB 2; Length 615;  
Best Local Similarity 75.0%; Pred. No. 2,15e+01;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
DB 163 IPYTVRK 170  
QY 1 IPYTVRK 8  
RESULT 13  
ENTRY S23541 #type complete  
TITLE hypothetical protein 125 - diatom (*Cylindrotheca fusiformis*)  
ORGANISM #formal\_name *Cylindrotheca fusiformis*  
DATE 03-May-1994 #sequence\_revision 20-Feb-1995 #text\_change  
09-Sep-1997  
ACCESSIONS S23541  
#authors Hildebrand, M.; Hasegawa, P.; Ord, R.W.; Thorpe, V.S.; Glass,  
C.A.; Volcani, B.E.  
#journal Plant Mol. Biol. (1992) 19:759-770  
#title Nucleotide sequence of diatom plasmids: identification of

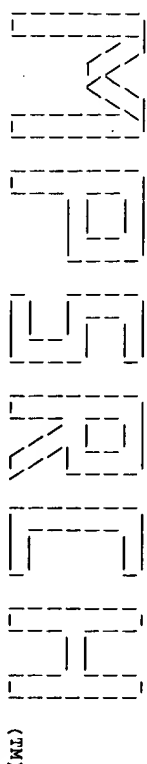
open reading frames with similarity to site-specific  
recombinases.  
#cross-references M1000147; GB:97426617  
#accession S23541  
#molecule-type DNA  
#residues 1-125 ##label HTL  
#cross-references EMBL:X64303; NID:g17994; PID:g18000  
GENETICS  
#gene plasmid  
#length 125 #molecular-weight 14573 #checksum 5702  
SUMMARY  
Query Match 67.1%; Score 49; DB 2; Length 125;  
Best Local Similarity 75.0%; Pred. No. 3,24e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
DB 64 PPIYVRL 71  
QY 2 PPIYVRL 9  
RESULT 14  
ENTRY E64506 #type complete  
TITLE adenine phosphoribosyltransferase (EC 2.4.2.7) -  
Methanococcus jannaschii  
ORGANISM #formal\_name *Methanococcus jannaschii*  
DATE 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change  
10-Oct-1997  
ACCESSIONS E64506  
#authors Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,  
R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,  
R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;  
Tomblin, J.F.; Adams, M.D.; Reisch, C.I.; Overbeek, R.;  
Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;  
Scott, J.L.; Geoghegan, N.S.M.; Feldman, J.F.; Fuhmann,  
J.L.; Nguyen, D.; Otterback, T.R.; Kelley, J.M.; Peterson,  
J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,  
K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,  
H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
#journal Science (1996) 273:1058-1073  
#title Complete genome sequence of the methanogenic archaeon,  
*Methanococcus jannaschii*.  
#cross-references M1000147; GB:97426617  
#accession E64506  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
#molecule-type DNA  
#residues 1-183 ##label BUL  
#cross-references GB:U67606; GB:L77117; NID:g1592235; PID:g1592237;  
TIGR:M1655; PID:g1511608  
GENETICS  
#map\_position FOR1637715-1638266  
#start-codon TTG  
KEYWORDS glycosyltransferase; pentosyltransferase  
SUMMARY #length 183 #molecular-weight 20218 #checksum 2893  
Query Match 67.1%; Score 49; DB 2; Length 183;  
Best Local Similarity 75.0%; Pred. No. 3,24e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
DB 76 IPYVIMR 83  
QY 1 IPYVIMR 8  
RESULT 15  
ENTRY C69432 #type complete  
TITLE hypothetical protein AFI460 - *Archaeoglobus fulgidus*  
ORGANISM #formal\_name *Archaeoglobus fulgidus*  
DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change  
05-Dec-1997  
ACCESSIONS C69432  
#authors  
#journal  
#title

```
#authors      Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
               K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
               Peterson, J.D.; Richardson, D.L.; Kervase, A.R.; Graham,
               D.E.; Kyriakides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
               Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
               Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
               Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
               Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
               J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
               T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
               D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
               Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
               C.R.; Venter, J.C.
#journal      Nature (1997) 390:364-370
#title        The complete genome sequence of the hyperthermophilic,
               sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MIMD:98049343
#accession    C69432
#status       preliminary; nucleic acid sequence not shown;
               translation not shown
##molecule_type DNA
##residues     1-227 ##label KLF
##cross-references GB:AE000782; TIGR:AF1460
SUMMARY       #length 227 #molecular-weight 26265 #checksum 1660

Query Match      67.1%; Score 49; DB 2; Length 227;
Best Local Similarity 55.6%; Pred. No. 3.24e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 219 IAYRVLRKL 227
      I : I : : : : :
QY      1 IYPYIVRKL 9

Search completed: Fri Sep 11 13:06:37 1998
Job time : 20 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*  
MUSCH.PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Sep 11 13:06:54 1998; MasPar time 2.35 Seconds  
Tabular output not generated. 96.097 Million cell updates/sec

Title: >US-08-452-843-8  
Description: (1-9) from US08452843.pep  
Perfect Score: 73  
Sequence: 1 IPIPIVRL 9

Scoring table: PAM 150  
Gap 15

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-Prot35  
1:swiss1

Statistics: Mean 25.074; Variance 27.458; scale 0.913

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID           | Description            | Pred. No. |
|------------|-------|-------------|--------|--------------|------------------------|-----------|
| 1          | 55    | 75.3        | 1679   | 1 Y109_YEAST | HYPOTHETICAL 195.1 KD  | 3.97e+01  |
| 2          | 53    | 72.6        | 209    | 1 RAS_GEOCY  | RAS-LIKE PROTEIN.      | 1.13e+00  |
| 3          | 53    | 72.6        | 487    | 1 YGJE_ECOLI | HYPOTHETICAL 52.9 KD P | 1.13e+00  |
| 4          | 52    | 71.2        | 880    | 1 RPA1_SULAC | DNA-DIRECTED RNA POLYM | 1.87e+00  |
| 5          | 51    | 69.9        | 340    | 1 RRC2_SCHPO | PROBABLE ACTIVATOR 1.4 | 3.10e+00  |
| 6          | 51    | 69.9        | 607    | 1 G6PI_TRYB  | GLUCOSE-6-PHOSPHATE IS | 3.10e+00  |
| 7          | 51    | 69.9        | 824    | 1 DPOL_METVO | DNA POLYMERASE (EC 2.7 | 5.09e+00  |
| 8          | 50    | 68.5        | 356    | 1 OXDA_TRIVR | D-AMINO ACID OXIDASE ( | 5.09e+00  |
| 9          | 50    | 68.5        | 615    | 1 SECD_ECOLI | PROTEIN-EXPORT MEMBRAN | 8.28e+00  |
| 10         | 49    | 67.1        | 183    | 1 APT_METJA  | ADENINE PHOSPHORIBOSYL | 8.28e+00  |
| 11         | 49    | 67.1        | 342    | 1 ARG1_ARATH | ARGINASE (EC 3.5.3.1). | 8.28e+00  |
| 12         | 49    | 67.1        | 1061   | 1 RNE_ECOLI  | RIBONUCLEASE E (EC 3.1 | 8.28e+00  |
| 13         | 49    | 67.1        | 1676   | 1 COS_HUMAN  | COMPLEMENT C5 PRECURSO | 8.28e+00  |
| 14         | 49    | 67.1        | 1680   | 1 COS_MOUSE  | COMPLEMENT C5 PRECURSO | 8.28e+00  |
| 15         | 48    | 65.8        | 100    | 1 YOG4_CAEEL | HYPOTHETICAL 11.2 KD P | 1.34e+01  |
| 16         | 48    | 65.8        | 305    | 1 LIGD_PSEPA | C ALPHA-DEHYDROGENASE  | 1.34e+01  |
| 17         | 48    | 65.8        | 315    | 1 HTRB_HAEIN | LIPID A BIOSYNTHESIS I | 1.34e+01  |
| 18         | 48    | 65.8        | 326    | 1 GLN2_BRJVA | GLUTAMINE SYNTHETASE I | 1.34e+01  |
| 19         | 48    | 65.8        | 356    | 1 YDGC_SCHPO | HYPOTHETICAL 41.3 KD P | 1.34e+01  |
| 20         | 48    | 65.8        | 372    | 1 DP3B_CAUCR | DNA POLYMERASE III, BE | 1.34e+01  |
| 21         | 48    | 65.8        | 377    | 1 TRAY_BACST | PUTATIVE TRANSPOSASE F | 1.34e+01  |
| 22         | 48    | 65.8        | 407    | 1 CPXD_AGR16 | CYCLOHROM PA50-PIN2,   | 1.34e+01  |
| 23         | 48    | 65.8        | 540    | 1 TH16_YEAST | THIAMIN BIOSYNTHETIC B | 1.34e+01  |

|    |    |      |      |              |                         |          |
|----|----|------|------|--------------|-------------------------|----------|
| 24 | 48 | 65.8 | 617  | 1 VGF8_RAT   | VGF8A PROTEIN PRECURSOR | 1.34e+01 |
| 25 | 48 | 65.8 | 762  | 1 TRPG_METOR | ANTHRANILATE SYNTHASE   | 1.34e+01 |
| 26 | 47 | 64.4 | 238  | 1 Y106_METJA | HYPOTHETICAL PROTEIN M  | 2.14e+01 |
| 27 | 47 | 64.4 | 268  | 1 YWDE_BACSU | HYPOTHETICAL 30.6 KD P  | 2.14e+01 |
| 28 | 47 | 64.4 | 323  | 1 YMC2_SCHPO | HYPOTHETICAL COXI INTR  | 2.14e+01 |
| 29 | 47 | 64.4 | 329  | 1 I329_ASFB7 | LATE PROTEIN I329L PRE  | 2.14e+01 |
| 30 | 47 | 64.4 | 329  | 1 GLN2_BRJVA | GLUTAMINE SYNTHETASE I  | 2.14e+01 |
| 31 | 47 | 64.4 | 381  | 1 Y931_METJA | HYPOTHETICAL PROTEIN M  | 2.14e+01 |
| 32 | 47 | 64.4 | 472  | 1 YAE3_SCHPO | HYPOTHETICAL 54.3 KD P  | 2.14e+01 |
| 33 | 47 | 64.4 | 682  | 1 GR78_YEAST | 78 KD GLUCOSE REGULATE  | 2.14e+01 |
| 34 | 47 | 64.4 | 1056 | 1 YNM2_YEAST | HYPOTHETICAL 119.3 KD   | 2.14e+01 |
| 35 | 47 | 64.4 | 1868 | 1 YHD0_YEAST | HYPOTHETICAL 210.4 KD   | 2.14e+01 |
| 36 | 46 | 63.0 | 164  | 1 YCEL_YEAST | HYPOTHETICAL 19.1 KD P  | 3.40e+01 |
| 37 | 46 | 63.0 | 206  | 1 PYRE_BACCL | OROTATE PHOSPHORIBOSYL  | 3.40e+01 |
| 38 | 46 | 63.0 | 245  | 1 OMCW_BOVIN | ONCOSTATIN M PRECURSOR  | 3.40e+01 |
| 39 | 46 | 63.0 | 409  | 1 YADE_ECOLI | HYPOTHETICAL 46.3 KD P  | 3.40e+01 |
| 40 | 46 | 63.0 | 581  | 1 YNM5_YEAST | HYPOTHETICAL 67.4 KD P  | 3.40e+01 |
| 41 | 46 | 63.0 | 662  | 1 PMT7_YEAST | DOLICHYL-PHOSPHATE-MAN  | 3.40e+01 |
| 42 | 46 | 63.0 | 752  | 1 TRPG_PENCH | ANTHRANILATE SYNTHASE   | 3.40e+01 |
| 43 | 46 | 63.0 | 768  | 1 TRPG_ASPAW | ANTHRANILATE SYNTHASE   | 3.40e+01 |
| 44 | 46 | 63.0 | 842  | 1 YJ67_YEAST | HYPOTHETICAL 99.3 KD P  | 3.40e+01 |
| 45 | 46 | 63.0 | 918  | 1 YNE6_CAEEL | HYPOTHETICAL 104.2 KD   | 3.40e+01 |

# ALIGNMENTS

| RESULT   | ID  | Y109_YEAST | STANDARD | PRT     | 1679 AA. |
|--|---|------------|----------|---------|----------|
| AC   | P40457  |            |          |         |          |
| DT   | 01-FEB-1995 (REL. 31, CREATED)                                      |            |          |         |          |
| DT   | 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)                         |            |          |         |          |
| DT   | 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)                       |            |          |         |          |
| DE   | HYPOTHETICAL 195.1 KD PROTEIN IN DNA43-UBI1 INTERGENIC REGION.      |            |          |         |          |
| GN   | Y11149C.  |            |          |         |          |
| OS   | SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).                           |            |          |         |          |
| OC   | EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCETES.                      |            |          |         |          |
| RN   | [1]   |            |          |         |          |
| RP   | SEQUENCE FROM N.A.  |            |          |         |          |
| RC   | STRAIN-5288C / AB972;   |            |          |         |          |
| RA   | BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,        |            |          |         |          |
| RA   | CHURCHER C.M., CONNOR R., COSEY T., DEAR S., DEVLIN K., FRASER A.,  |            |          |         |          |
| RA   | GENTLES S., HAWLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M., |            |          |         |          |
| RA   | LOUIS E., LYE G., MOUE S., MOUE T., ODELL C., PEARSON D.,           |            |          |         |          |
| RA   | RAUDREMAN M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,          |            |          |         |          |
| RA   | WALSH S.V., WHITEHEAD S.;   |            |          |         |          |
| RL   | SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.                |            |          |         |          |
| DR   | EMBL; 247047; G763197; -  |            |          |         |          |
| DR   | EMBL; 238059; G557774; -  |            |          |         |          |
| DR   | PIR; S48385; S48385.  |            |          |         |          |
| KW   | HYPOTHETICAL PROTEIN.   |            |          |         |          |
| SQ   | SEQUENCE 1679 AA; 195141 MW; 5897CD94 CRC32;                        |            |          |         |          |
| Query Match  |   |            |          |         |          |
| Best Local Similarity 75.3% Score 55; DB 1; Length 1679;   |   |            |          |         |          |
| Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0; |   |            |          |         |          |
| DB   | 19 YIPIVRL 27   |            |          |         |          |
| QY   | 1 IPIPIVRL 9  |            |          |         |          |
| RESULT 2   |   |            |          |         |          |
| ID   | RAS_GEOCY   | STANDARD;  | PRT;     | 209 AA. |          |
| AC   | P24498;   |            |          |         |          |
| DT   | 01-MAR-1992 (REL. 21, CREATED)                                      |            |          |         |          |
| DT   | 01-FEB-1992 (REL. 21, LAST SEQUENCE UPDATE)                         |            |          |         |          |
| DT   | 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)                       |            |          |         |          |
| DE   | RAS-LIKE PROTEIN.   |            |          |         |          |
| OS   | GEODIA CYDONIUM (SPONGE).   |            |          |         |          |
| OC   | EUKARYOTA; METAZOA; PORIFERA.                                       |            |          |         |          |
| RP   | SEQUENCE FROM N.A.  |            |          |         |          |
| RX   | MEDLINE; 91006138.  |            |          |         |          |

RA ROBITZKI A., SCHROEDER H.C., UGAROVIC D., KUCHINO Y., KURELEC B.,  
 RA GAMLIN V., MEILLER W.E.G.;  
 RL EUR. J. BIOCHEM. 192:499-506(1990).  
 CC -1- FUNCTION: THIS PROTEIN IS ACTIVATED BY THE INSULIN/INSULIN  
 CC (INSULIN-LIKE)-RECEPTOR SYSTEM. THIS TRANSITION ENABLES THE RAS  
 CC PROTEIN TO INTERACT WITH THE LECTIN-RECEPTOR/LECTIN COMPLEX. A  
 CC PROCESS WHICH ULTIMATELY LEAD TO AN INITIATION OF AN INTRA-  
 CC CELLULAR SIGNAL-TRANSDUCTION CHAIN.  
 CC -1- PTM: PHOSPHORYLATED IN THE PRESENCE OF INSULIN.  
 DR EMBL: M30929; E51958; ALT-SEQ.  
 DR PIR: S13179; S13179.  
 DR HSSP: P01112; 1PLT.  
 KW GTP-BINDING; PRENYLATION; LIPOPROTEIN; PHOSPHORYLATION.  
 FT NP-BIND 10 17 GTP (BY SIMILARITY).  
 FT NP-BIND 79 83 GTP (BY SIMILARITY).  
 FT NP-BIND 140 143 GTP (BY SIMILARITY).  
 FT DOMAIN 55 63 EFFECTOR REGION (BY SIMILARITY).  
 FT MOD\_RES 58 58 PHOSPHORYLATION (POTENTIAL).  
 FT LIPID 206 206 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 209 AA; 23854 MW; E07739EF CRC32;

Query Match 72.6%; Score 53; DB 1; Length 209;  
 Best Local Similarity 66.7%; Pred. No. 1.13e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 178 IPIVRL 186  
 1 IPIVRL 9

RESULT 3  
 ID YGJE\_ECOLI STANDARD; PRT; 487 AA.  
 AC P39414; Q46870;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHEICAL 52.9 KD PROTEIN IN TDB-RPSU INTERGENIC REGION.  
 GN YGJE.  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA BLATTNER F.R., PUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBD DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE: 87248073.  
 RA NESIN M., LUPSKI J.R., SVEC P., GODSON G.N.;  
 RL GENE 51:149-161(1987).  
 RN [3]  
 RP IDENTIFICATION.  
 RA MEDLINE: 95075659.  
 RA BORODOVSKY M., RUDD K.E., KOONIN E.V.;  
 RL NUCLEIC ACIDS RES. 22:4756-4767(1994).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE MDC/P/PHO87 FAMILY OF TRANSPORTERS.  
 CC SODIUM SUBFAMILY.  
 CC -1- CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS  
 CC IN POSITIONS 35, 51, 132, 245, 268 AND 443.  
 DR EMBL: U28379; G882586;  
 DR EMBL: AE000388; G1789444;  
 DR EMBL: M16194; -; NOT ANNOTATED\_CDS.  
 DR ECGENE: EG13393; YGJE.  
 KW HYPOTHEICAL PROTEIN; TRANSMEMBRANE; INNER MEMBRANE; TRANSPORT.  
 FT TRANSMEM 10 30 POTENTIAL.  
 FT TRANSMEM 33 53 POTENTIAL.  
 FT TRANSMEM 54 74 POTENTIAL.  
 FT TRANSMEM 93 113 POTENTIAL.  
 FT TRANSMEM 137 157 POTENTIAL.  
 FT TRANSMEM 189 209 POTENTIAL.

FT TRANSMEM 236 256 POTENTIAL.  
 FT TRANSMEM 292 312 POTENTIAL.  
 FT TRANSMEM 313 333 POTENTIAL.  
 FT TRANSMEM 340 360 POTENTIAL.  
 FT TRANSMEM 370 390 POTENTIAL.  
 FT TRANSMEM 393 413 POTENTIAL.  
 FT TRANSMEM 418 438 POTENTIAL.  
 FT TRANSMEM 465 485 POTENTIAL.  
 FT TRANSMEM 404 404 L -> P (IN REF. 2).  
 FT TRANSMEM 457 457 A -> T (IN REF. 2).  
 FT TRANSMEM 457 457 A -> T (IN REF. 2).  
 SQ SEQUENCE 487 AA; 52906 MW; EB673FE9 CRC32;

Query Match 72.6%; Score 53; DB 1; Length 487;  
 Best Local Similarity 66.7%; Pred. No. 1.13e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 165 IPIVRL 173  
 1 IPIVRL 9

RESULT 4  
 ID RPA1\_STLAC STANDARD; PRT; 880 AA.  
 AC P1512;  
 DT 01-OCT-1989 (REL. 12, CREATED)  
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE DNA-DIRECTED RNA POLYMERASE SUBUNIT A' (EC 2.7.7.6).  
 GN RPOA1 OR RPOA.  
 OS SULFOLOBUS ACIDOCALDARIUS.  
 OC ARCHAEABACTERIA; CRENARCHAEOTA; SULFOLOBALES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DSK 639;  
 RA MEDLINE: 89315197.  
 RA PUEHLER G., LOTSPRECH F., ZILLIG W.;  
 RL NUCLEIC ACIDS RES. 17:4517-4534(1989).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -1- COFACTOR: ZINC.  
 CC -1- SUBUNIT: S.ACIDOCALDARIUS RNAP IS COMPOSED OF 13 SUBUNITS.  
 CC -1- SIMILARITY: THE COMBINED A'+A' SUBUNITS CORRESPOND TO THE A  
 CC SUBUNITS OF EUKARYOTIC RNA POLYMERASES I, II AND III AND TO THE  
 CC EUBACTERIAL BETA' SUBUNIT.  
 DR EMBL: X14818; G46670;  
 DR PIR: S04717; S04717.  
 DR TRANSCRIPTION: DNA-DIRECTED RNA POLYMERASE. ZINC.  
 KW ZN\_FING 58 101 POTENTIAL.  
 FT ZN\_FING 58 101 POTENTIAL.  
 SQ SEQUENCE 880 AA; 99790 MW; 665B33F9 CRC32;

Query Match 71.2%; Score 52; DB 1; Length 880;  
 Best Local Similarity 66.7%; Pred. No. 1.87e+00;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 339 IPIVRL 347  
 1 IPIVRL 9

RESULT 5  
 ID RFG2\_SCHPO STANDARD; PRT; 340 AA.  
 AC Q09843;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE PROBABLE ACTIVATOR 1 41 KD SUBUNIT (REPLICATION FACTOR C 41 KD  
 DE SUBUNIT).  
 GN SPAC23D3.02.  
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
 GN EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.



RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972;  
 RA NIBLET D., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;  
 RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: THE ELONGATION OF PRIMED DNA TEMPLATES BY DNA POLYMERASE  
 CC DELTA AND EPSILON REQUIRES THE ACTION OF THE ACCESSORY PROTEINS  
 CC PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) AND ACTIVATOR 1. THE  
 CC 41 KD SUBUNIT BINDS ATP AND TO SINGLE-STRANDED DNA  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE ACTIVATOR 1 36 TO 40 KD SUBUNITS  
 CC FAMILY.  
 DR EMBL: 264354; E205682;  
 KM HYPOHETHEICAL PROTEIN; DNA REPLICATION; ATP-BINDING; NUCLEAR PROTEIN;  
 KW DNA-BINDING.  
 FT NP-BIND 59 66 ATP (POTENTIAL).  
 SQ SEQUENCE 340 AA; 37876 MW; FB518443 CRC32;  
  
 Db 249 VPYNIIRSL 257  
 QY 1 IYPIVIRKL 9  
  
 RESULT 6  
 ID 66PI-TRYB STANDARD; PRT: 607 AA.  
 AC P13377;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)  
 DE GLUCOSE-6-PHOSPHATE ISOMERASE, GLYCOSOMAL (GPI) (EC 5.3.1.9)  
 DE (PHOSPHOGLUCOSE ISOMERASE) (PGI) (PHOSPHOHEXOSE ISOMERASE) (PHI).  
 GN PGI.  
 OS TRYPAOSOMA BRUCEI BRUCEI.  
 OC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA; KINETOPLASTIDA;  
 OC TRYPANOSOMATIDAE.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-427;  
 RX MEDLINE: 90005496.  
 RA MARCHAND M., KOOSTRA U., WIERENGA R.K., LAMBEIR A.M., VAN BEUDEN J.,  
 RA OPPERDUSE F.R., MICHELIS P.A.M.;  
 RL EUR. J. BIOCHEM. 184:455-464(1989).  
 CC -1- CATALYTIC ACTIVITY: GLUCOSE 6-PHOSPHATE = FRUCTOSE 6-PHOSPHATE.  
 CC -1- PATHWAY: INVOLVED IN GLYCOLYSIS AND IN GLUCONEOGENESIS.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: GLYCOSOMAL.  
 CC -1- SIMILARITY: HIGH WITH GPI FROM OTHER SPECIES.  
 DR EMBL: X15540; G10487;  
 DR PIR: S06113; NUTB.  
 DR PROSITE: PS00174; P-GLUCOSE ISOMERASE 2; 1.  
 DR PROSITE: PS00342; MICROBODIES-CTER; 1.  
 DR PROSITE: PS00765; P-GLUCOSE ISOMERASE 1; 1.  
 KW GLUCONEOGENESIS; GLYCOLYSIS; ISOMERASE; GLYCOSOME.  
 FT SITE 605 607 MICROBODY TARGETING SIGNAL (POTENTIAL).  
 SQ SEQUENCE 607 AA; 67518 MW; EF35C43 CRC32;  
  
 Query Match 69.9%; Score 51; DB 1; Length 607;  
 Best Local Similarity 55.6%; Pred. No. 3.10e+00;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE DNA POLYMERASE (EC 2.7.7.7).  
 OS METHANOCOCCUS VOLTAE.  
 OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE: 95014087.  
 RA KONISKY J., PAULE S.M., CARINATO M.E., KANSY J.W.;  
 RL J. BACTERIOL. 176:6402-6403(1994).  
 CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -  
 CC N PYROPHOSPHATE + DNA(N).  
 CC -1- SIMILARITY: BELONGS TO FAMILY B OF DNA POLYMERASES.  
 DR EMBL: L33366; G495654;  
 DR PROSITE: PS00116; DNA\_POLYMERASE\_B; 1.  
 KW DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION; DNA-BINDING.  
 SQ SEQUENCE 824 AA; 96754 MW; 94579170 CRC32;  
  
 Db 308 YPIARL 314  
 QY 3 YPIVIRKL 9  
  
 RESULT 8  
 ID OXDA-TRIVR STANDARD; PRT: 356 AA.  
 AC O99042;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE D-AMINO ACID OXIDASE (EC 1.4.3.3) (DAMO) (DAO) (DAAO).  
 OS TRICHOPOPSIS VARIABILIS.  
 OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CBS 4095;  
 RA GONZALEZ F.;  
 RL THESES (1996), UNIVERSIDAD DE SALAMANCA, SPAIN.  
 CC -1- CATALYTIC ACTIVITY: A D-AMINO ACID + H(2)O + O(2) = A 2-OXO-ACID +  
 CC NH(3) + H(2)O(2).  
 CC -1- COFACTOR: FAD FLAVOPROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.  
 DR EMBL: 250019; E187982;  
 DR PROSITE: PS00677; DAO; 1.  
 DR OXIDOREDUCTASE; FLAVOPROTEIN; FAD.  
 KM NP-BIND 4 18  
 FT ACT\_SITE 243 243 BY SIMILARITY.  
 FT ACT\_SITE 324 324 BY SIMILARITY.  
 SQ SEQUENCE 356 AA; 39301 MW; BA069642 CRC32;  
  
 Query Match 68.5%; Score 50; DB 1; Length 356;  
 Best Local Similarity 55.6%; Pred. No. 5.09e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 31 IYEVTRRL 39  
 QY 1 IYPIVIRKL 9  
  
 RESULT 7  
 ID DPOL-METVO STANDARD; PRT: 824 AA.  
 AC P52025;  
  
 Query Match 69.9%; Score 51; DB 1; Length 607;  
 Best Local Similarity 55.6%; Pred. No. 3.10e+00;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

ID SECD-ECOLI STANDARD; PRT: 615 AA.  
 AC P19673; P77531; P72348;  
 DT 01-FEB-1991 (REL. 17, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE PROTEIN-EXPORT MEMBRANE PROTEIN SECD.  
 GN SECD.  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;

OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA MEDLINE: 9106014.  
 RA GARDEL C., JOHNSON K., JACO A., BECKWITH J.;  
 RL EMBO J. 9:3209-3216(1990).  
 RN [2]  
 RP ERRATUM.  
 RA MEDLINE: 91065354.  
 RA GARDEL C., JOHNSON K., JACO A., BECKWITH J.;  
 RL EMBO J. 9:4205-4206(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,  
 RA DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,  
 RA LER H., LIN D., NAKATH A., OEPNER P., SCHRAM S., DAVIS R.W.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [5]  
 RP SEQUENCE OF 1-76 FROM N.A.  
 RX MEDLINE: 94131960.  
 RA POSTIANO K.J., BECKWITH J.;  
 RL J. BACTERIOL. 176:804-814(1994).  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.  
 CC -1- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECA-F & SECY) THAT  
 CC COMPRISE THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
 CC -1- SIMILARITY: IN THE C-TERMINAL, WITH SECF.  
 DR EMBL: X56175; G581230; -;  
 DR EMBL: AE000147; G1786609; -;  
 DR EMBL: U82664; G1773092; -;  
 DR EMBL: S68715; G545175; -;  
 DR PIR: J00696; J00696;  
 DR PIR: S12301; S12301;  
 DR ECGENE: EG10938; SECD.  
 KW PROTEIN TRANSPORT; TRANSLOCATION; TRANSMEMBRANE; INNER MEMBRANE.  
 FT TRANSMEM 10 30 POTENTIAL.  
 FT TRANSMEM 452 472 POTENTIAL.  
 FT TRANSMEM 504 524 POTENTIAL.  
 FT TRANSMEM 564 584 POTENTIAL.  
 FT CONFLICT 78 78 F -> S (IN REF. 1).  
 FT CONFLICT 155 155 R -> A (IN REF. 1).  
 SQ SEQUENCE 615 AA; 66632 MW; 9943E19B CRC32;  
 Query Match 68.5%; Score 50; DB 1; Length 615;  
 Best Local Similarity 75.0%; Pred. No. 5.09e+00;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 163 IPYTVRK 170  
 QY 1 IPYTVRK 8

RESULT 10  
 ID APT.METUA STANDARD; PRT; 183 AA.  
 AC 059049;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ADENINE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.7) (APRT).  
 GN APT OR MJ1655.  
 OS METHANOCOCCUS JANNASCHII.  
 OC ARCHAEABACTERIA; EUVYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96337999.  
 RA BOLT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,

RA KERLAUGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
 RA OVERBERG R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,  
 RA SCOTT J.L., GEORGESEN N.S.M., WEIDMAN J.F., FUHRMAN J.L., NGUYEN D.,  
 RA UTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,  
 RA KLEIN H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;  
 RL SCIENCE 273:1058-1073(1996).  
 CC -1- FUNCTION: CATALYZES A SALVAGE REACTION RESULTING IN THE FORMATION  
 CC OF AMP, THAT IS ENERGETICALLY LESS COSTLY THAN DE NOVO SYNTHESIS.  
 CC -1- CATALYTIC ACTIVITY: AMP + PYROPHOSPHATE = ADENINE + 5-PHOSPHO-  
 CC ALPHA-D-RIBOSE 1-DIPHOSPHATE.  
 CC -1- PATHWAY: PURINE SALVAGE.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).  
 CC -1- SIMILARITY: PARTIAL WITH OTHER PHOSPHORIBOSYLTRANSFERASE ALSO  
 CC INVOLVED IN BIOSYNTHESIS OR SALVAGE OF PURINES OR PYRIMIDINES.  
 DR EMBL: U67606; G159237; -;  
 DR PROSITE: PS00103; PUR\_PYR\_PR\_TRANSFER; 1.  
 DR TIGR: MJ1655; -;  
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; PURINE SALVAGE.  
 SQ SEQUENCE 183 AA; 20218 MW; E968EC1 CRC32;  
 Query Match 67.1%; Score 49; DB 1; Length 183;  
 Best Local Similarity 75.0%; Pred. No. 8.28e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 76 IPYTVRK 83  
 QY 1 IPYTVRK 8

RESULT 11  
 ID ANGLARATH STANDARD; PRT; 342 AA.  
 AC P46637;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ARGININASE (EC 3.5.3.1).  
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;  
 OC CAMPARALES; CERCIFERAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV LANDSBERG ERRECTA;  
 RA KROMPELMAN P.M., FREYERMUTH S.K., CANNON J.F., FINK G.R.,  
 RA POLACCO J.C.;  
 RL PLANT PHYSIOL. 107:1479-1480(1995).  
 CC -1- CATALYTIC ACTIVITY: L-ARGININE + H(2)O = L-ORNITHINE + UREA.  
 CC -1- COFACTOR: MN(2+) (BY SIMILARITY).  
 CC -1- PATHWAY: FIRST STEP IN ARGININE DEGRADATION.  
 CC -1- SIMILARITY: BELONGS TO THE ARGININASE FAMILY.  
 DR EMBL: U15019; G602422; -;  
 DR PROSITE: PS00147; ARGININASE\_1; 1.  
 DR PROSITE: PS00148; ARGININASE\_2; 1.  
 DR PROSITE: PS01053; ARGININASE\_3; 1.  
 KW HYDROLASE; ARGININE METABOLISM; MANGANESE.  
 FT METAL 161 161 MANGANESE 1 (BY SIMILARITY).  
 FT METAL 185 185 MANGANESE 1 AND 2 (BY SIMILARITY).  
 FT METAL 187 187 MANGANESE 2 (BY SIMILARITY).  
 FT METAL 189 189 MANGANESE 1 (BY SIMILARITY).  
 FT METAL 270 270 MANGANESE 1 AND 2 (BY SIMILARITY).  
 FT METAL 272 272 MANGANESE 2 (BY SIMILARITY).  
 SQ SEQUENCE 342 AA; 37344 MW; 9640021A CRC32;  
 Query Match 67.1%; Score 49; DB 1; Length 342;  
 Best Local Similarity 71.4%; Pred. No. 8.28e+00;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 163 ISYVVR 169  
 QY 1 IPYTVRK 7

RESULT 12  
ID RNE\_ECOLI STANDARD: PRT: 1061 AA.  
AC P21513;  
DT 01-MAY-1991 (REL. 18, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE RIBONUCLEASE E (EC 3.1.4.-) (RNASE E).  
GN RNE OR AMS OR HMPI.  
OS ESCHERICHIA COLI.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
ENTEROBACTERIACEAE.  
[1]  
[1] SEQUENCE OF 1-1025 FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE: 93078265.  
RA CASAREBOLA S., JACO A., LAOUDJ D., MCGURK G., MARGARSON S.,  
RA TEMPESTE M., NORRIS V., HOLLAND I.B.;  
RL J. MOL. BIOL. 228:30-40(1992).  
[2]  
[2] SEQUENCE OF 1-844 FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE: 91131576.  
RA CLAVERIE-MARTIN F., DIAZ-TORRES M., YANCEY S.D., KUSHNER S.R.;  
RL J. BIOL. CHEM. 266:2843-2851(1991).  
[3]  
[3] PARTIAL SEQUENCE FROM N.A., AND SEQUENCE OF 1-27.  
RC STRAIN-K12;  
RX MEDLINE: 91187608.  
RA CHAHDHAN A.K., MICZAK A., TARASEVICIENE L., APIRION D.;  
RL NUCLEIC ACIDS RES. 19:125-129(1991).  
[4]  
[4] SEQUENCE OF 844-1061 FROM N.A., AND CHARACTERIZATION.  
RC STRAIN-K12;  
RX MEDLINE: 94022304.  
RA CORMACK R.S., GENEREAUX J.L., MACKIE G.A.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 90:9006-9010(1993).  
[5]  
[5] FUNCTION: THIS PROTEIN MATURES 35 RNA FROM ITS PRECURSORS FROM  
ALL THE RNA GENES. IT ALSO CLEAVES RNA I, A MOLECULE THAT  
CONTROLS THE REPLICATION OF COLEI PLASMID DNA. IT IS THE MAJOR  
ENRORIBONUCLEASE PARTICIPATING IN MRNA TURNOVER IN E.COLI.  
[6]  
[6] SUBUNIT: ORGANISED INTO A STRUCTURE (PROCESSOME OR RNA  
DEGRADOSOME) CONTAINING A NUMBER OF RNA-PROCESSING ENZYMES.  
[7]  
[7] SUBCELLULAR LOCATION: CYTOPLASMIC.  
[8]  
[8] CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1003  
ONWARD AND IS SHORTER (1025 AA) DUE TO A FRAMESHIFT.  
[9]  
[9] CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE C-TERMINUS  
AND IS SHORTER (815 AA) DUE TO A FRAMESHIFT.  
[10]  
[10] CAUTION: REF.3 SEQUENCE WAS ALSO INCORRECT IN MANY POSITIONS DUE  
TO FRAMESHIFTS.  
DR EMBL: X67470; G49116; ALT\_FRAME.  
DR EMBL: M62747; G145273; ALT\_FRAME.  
DR EMBL: X54309; G42773; ALT\_FRAME.  
DR EMBL: L23942; G397760; -.  
DR PIR: JG0009; JG0009.  
DR PIR: A23747; A23747.  
DR PIR: S25116; S25116.  
DR PIR: S27311; S27311.  
DR ECGENE: EG10859; RNE.  
KW HYDROLASE; NUCLEASE; ENDONUCLEASE; RNA-BINDING.  
FT CONFLICT 390 390 Q -> H (IN REF. 2).  
FT CONFLICT 564 564 R -> A (IN REF. 2).  
FT CONFLICT 784 784 K -> N (IN REF. 2).  
FT CONFLICT 838 838 A -> R (IN REF. 2).  
FT CONFLICT 905 905 P -> R (IN REF. 1).  
SQ SEQUENCE 1061 AA; 118301 MW; 3473778C CRC32;

Query Match 67.1%; Score 49; DB 1; Length 1061;  
Best Local Similarity 85.7%; Pred. No. 8.28e+00;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 844 INPIV 850  
| | | | |

QY 1 INPIV 7  
RESULT 13  
ID COS\_HUMAN STANDARD: PRT: 1676 AA.  
AC P01031;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE COMPLEMENT C5 PRECURSOR (CONTAINS: C5A ANAPHYLATOXIN).  
GN C5.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
[1]  
[1] SEQUENCE FROM N.A.  
RX MEDLINE: 91079575.  
RA HAYLAND D.L., HAYLAND J.C., FLEISCHER D.T., HUNT A., WETSEL R.A.;  
RL J. IMMUNOL. 146:362-368(1991).  
[2]  
[2] SEQUENCE OF 412-1676 FROM N.A.  
RX MEDLINE: 88209511.  
RA WETSEL R.A., LEMONS R.S., LEBEAU M.M., BARNUM S.R., NOACK D.,  
RA TACK B.F.;  
RL BIOCHEMISTRY 27:1474-1482(1988).  
[3]  
[3] SEQUENCE OF 412-902 FROM N.A.  
RX MEDLINE: 85130937.  
RA LUNDWALL A.B., WETSEL R.A., KRISTENSEN T., WHITEHEAD A.S.,  
RA WOODS D.E., OGDEN R.C., COLTEN H.R., TACK B.F.;  
RL J. BIOL. CHEM. 260:2108-2112(1985).  
[4]  
[4] SEQUENCE OF 678-751.  
RX MEDLINE: 79005687.  
RA FERNANDEZ H.N., HUGLI T.E.;  
RL J. BIOL. CHEM. 253:6955-6964(1978).  
[5]  
[5] SEQUENCE OF 678-751 FROM N.A.  
RX MEDLINE: 91144547.  
RA BORNISACK J.F., MOLLISON K.W., BURK A.M., ASHWORTH J.C., HILL H.R.;  
RL BIOCHEM. J. 273:635-640(1991).  
[6]  
[6] STRUCTURE BY NMR OF C5A.  
RX MEDLINE: 88309754.  
RA ZUIDERWEG E.R., MOLLISON K.W., HENKIN J., CARTER G.W.;  
RL BIOCHEMISTRY 27:3568-3580(1988).  
[7]  
[7] STRUCTURE BY NMR OF C5A.  
RX MEDLINE: 89207527.  
RA ZUIDERWEG E.R., NETTESHEIM D.G., MOLLISON K.W., CARTER G.W.;  
RL BIOCHEMISTRY 28:172-185(1989).  
[8]  
[8] STRUCTURE BY NMR OF C5A.  
RX MEDLINE: 89274164.  
RA ZUIDERWEG E.R., FESIK S.W.;  
RL BIOCHEMISTRY 28:2387-2391(1989).  
[9]  
[9] STRUCTURE BY NMR OF C5A.  
RX MEDLINE: 97160477.  
RA ZHANG X., BOYAR W., GALAKATOS N., GONNELLA N.C.;  
RL PROTEIN SCI. 6:65-72(1997).  
[10]  
[10] STRUCTURE BY NMR OF C5A.  
RX MEDLINE: 97332508.  
RA ZHANG X., BOYAR W., TOH M.J., WENNOGLE L., GONNELLA N.C.;  
RL PROTEINS 28:261-267(1997).  
[11]  
[11] FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE  
SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,  
INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE  
FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYtic  
COMPLEX IS ASSEMBLED.  
[12]  
[12] FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,  
C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT

INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).

-1- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN, RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA CHAIN).

-1- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.

-1- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.

CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 855 ONWARD DUE TO THE PRESENCE OF AN ALU REPEAT.

EMBL: M57729; G179983; -

EMBL: M65134; G179692; -

PIR: A40075; C5HU.

PIR: S15121; S15121.

PDB: 1KJ5; 15-MAY-97.

PDB: 1CFA; 17-SEP-97.

MIM: 120900; -

PROSITE: PS00477; ALPHA\_2\_MACROGLOBULIN; FALSE\_NEG.

PROSITE: PS01177; ANAPHYLATOXIN\_1; 1.

PROSITE: PS01178; ANAPHYLATOXIN\_2; 1.

KW COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; GLYCOPROTEIN; PLASMA; MEMBRANE ATTACK COMPLEX; CYTOLYSIS; INFLAMMATORY RESPONSE; SIGNAL; POLYMORPHISM; 3D-STRUCTURE.

FT SIGNAL 1 18 POTENTIAL.

FT CHAIN 19 673 COMPLEMENT C5 BETA CHAIN.

FT PROPEP 674 677

FT CHAIN 678 1676 COMPLEMENT C5 ALPHA CHAIN.

FT PEPTIDE 678 751 C5A ANAPHYLATOXIN.

FT CHAIN 752 1676 C5B (ALPHA').

FT DOMAIN 698 732 ANAPHYLATOXIN-LIKE.

FT DISULFID 698 724

FT DISULFID 699 731

FT DISULFID 711 732

FT CARBOHYD 741 741

FT CARBOHYD 911 911

FT CARBOHYD 1115 1115

FT CARBOHYD 1630 1630 POTENTIAL.

FT VARIANT 518 518 F -> S.

SO SEQUENCE 1676 AA; 188331 MW; 9D5C6E59 CRC32;

Query Match 67.1%; Score 49; DB 1; Length 1676;  
Best Local Similarity 71.4%; Pred. No. 8.28e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 829 IPIYSVR 835  
1 IPIYIVR 7

RESULT 14  
ID COS\_MOUSE STANDARD; PRT; 1680 AA.  
AC P06684;  
DT 01-JAN-1988 (REL. 06, CREATED)  
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE COMPLEMENT\_C5\_PRECURSOR (CONTAINS: C5A ANAPHYLATOXIN).  
GN C5 OR HC.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90153853.  
RA WETSEL R.A., FLEISCHER D.T., HAYLAND D.L.;  
RL J. BIOL. CHEM. 265:2435-2440(1990).  
RN [2]  
RP SEQUENCE OF 41-1680 FROM N.A.  
RX MEDLINE: 87185363.  
RA WETSEL R.A., OGATA R.T., TACK B.F.;

BIOCHEMISTRY 26:737-743(1987).

-1- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9, INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYtic COMPLEX IS ASSEMBLED.

-1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5, C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).

-1- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN, RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA CHAIN).

-1- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.

-1- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.

EMBL: M35525; G309124; -

EMBL: M35526; G309123; -

PIR: A27538; A27538.

PIR: A35530; A35530.

HSSP: P01032; 1C5A.

MCD: MGI:96031; HC.

PROSITE: PS00477; ALPHA\_2\_MACROGLOBULIN; FALSE\_NEG.

PROSITE: PS01177; ANAPHYLATOXIN\_1; 1.

PROSITE: PS01178; ANAPHYLATOXIN\_2; 1.

KW COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; GLYCOPROTEIN; PLASMA; MEMBRANE ATTACK COMPLEX; CYTOLYSIS; INFLAMMATORY RESPONSE; SIGNAL.

FT SIGNAL 1 18

FT CHAIN 19 1680

FT CHAIN 19 674 COMPLEMENT C5.

FT PROPEP 675 678 COMPLEMENT C5 BETA CHAIN.

FT CHAIN 679 1680

FT PEPTIDE 679 755 COMPLEMENT C5 ALPHA CHAIN.

FT CHAIN 756 1680 C5A ANAPHYLATOXIN.

FT DOMAIN 702 736 C5B (ALPHA').

FT DISULFID 702 728 ANAPHYLATOXIN-LIKE.

FT DISULFID 703 735 BY SIMILARITY.

FT DISULFID 715 736 BY SIMILARITY.

FT CARBOHYD 427 427 POTENTIAL.

FT CARBOHYD 915 915 POTENTIAL.

FT CARBOHYD 1119 1119 POTENTIAL.

FT CARBOHYD 1633 1633 POTENTIAL.

FT VARIANT 216 216 Y -> L (IN DEFECTIVE VARIANT C5D).

FT VARIANT 217 1680 MISSING (IN DEFECTIVE VARIANT C5D).

SO SEQUENCE 1680 AA; 188877 MW; AA17044B CRC32;

Query Match 67.1%; Score 49; DB 1; Length 1680;  
Best Local Similarity 71.4%; Pred. No. 8.28e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 833 IPIYSVR 839  
1 IPIYIVR 7

RESULT 15  
ID YOGA\_CAMEL STANDARD; PRT; 100 AA.  
AC P34613;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 11.2 KD PROTEIN ZK112.4 IN CHROMOSOME III.  
GN ZK112.4.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL NZ;

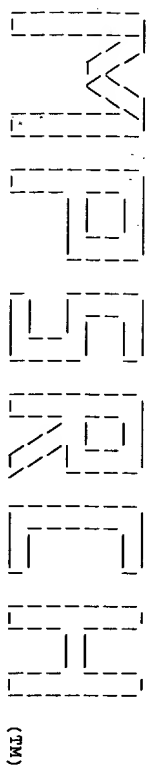
RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERRS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,  
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,  
 RA EUTTON L., GRADNER A., GREEN P., HAWKINS T., HILLIER L., JER M.,  
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSIER N.,  
 RA LATREUILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,  
 RA SUTSTON J., THIRRY-MING J., THOMAS K., VAUDIN M., VAUGHAN K.,  
 RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,  
 RA WOHLDMAN P.,  
 RL NATURE 368:32-38(1994).  
 DR EMBL: L14324; G289747; .  
 DR PIR: S44892; S44892.  
 DR WORMPEP; ZK112.4; CE00375.  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 100 AA; 11248 MW; 52F18207 CRC32;

Query Match 65.8%; Score 48; DB 1; Length 100;  
 Best Local Similarity 62.5%; Pred. NO. 1.34e+01;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 83 PPIVSRL 90  
 1:111:1  
 QY 2 PYPIVRKL 9

Search completed: Fri Sep 11 13:07:01 1998  
 Job time : 7 secs.

**THIS PAGE BLANK (USPTO)**



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Sep 11 13:07:20 1998; Maspar time 4.18 Seconds  
Tabular output not generated. 90.697 Million cell updates/sec

Title: >US-08-452-843-8  
Description: (1-9) from US08452843.pep  
Perfect Score: 73  
Sequence: 1 IIPPIYRKL 9

Scoring table: PAM 150  
Gap 15

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

sptrembl6  
1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_phc 6:sp\_organelle 7:sp\_phase 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp\_vertebrate  
13:sp\_unclassified

Statistics: Mean 24.375; Variance 30.315; scale 0.804

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|--------|-------|-------------|-----------|
| 1          | 59    | 80.8        | 410    | 9     | 033331      | 2.28e+01  |
| 2          | 53    | 72.6        | 52     | 9     | 033283      | 4.13e+00  |
| 3          | 53    | 72.6        | 1033   | 1     | P87115      | 6.56e+00  |
| 4          | 52    | 71.2        | 451    | 9     | P74054      | 1.04e+01  |
| 5          | 51    | 69.9        | 75     | 9     | 031936      | 1.04e+01  |
| 6          | 51    | 69.9        | 112    | 11    | Q84619      | 1.04e+01  |
| 7          | 51    | 69.9        | 133    | 10    | 009006      | 1.04e+01  |
| 8          | 51    | 69.9        | 133    | 10    | 009002      | 1.04e+01  |
| 9          | 51    | 69.9        | 197    | 9     | 058389      | 1.04e+01  |
| 10         | 51    | 69.9        | 766    | 9     | 051849      | 1.04e+01  |
| 11         | 50    | 68.5        | 129    | 11    | 036900      | 1.63e+01  |
| 12         | 50    | 68.5        | 130    | 11    | 036786      | 1.63e+01  |
| 13         | 50    | 68.5        | 280    | 3     | 020408      | 1.63e+01  |
| 14         | 50    | 68.5        | 357    | 9     | 006235      | 1.63e+01  |
| 15         | 50    | 68.5        | 491    | 11    | Q78998      | 1.63e+01  |
| 16         | 49    | 67.1        | 125    | 8     | Q39508      | 2.53e+01  |
| 17         | 49    | 67.1        | 227    | 9     | 028812      | 2.53e+01  |
| 18         | 49    | 67.1        | 396    | 9     | 006038      | 2.53e+01  |
| 19         | 49    | 67.1        | 537    | 9     | 046977      | 2.53e+01  |
| 20         | 49    | 67.1        | 598    | 3     | 016669      | 2.53e+01  |

|    |    |      |      |    |        |          |
|----|----|------|------|----|--------|----------|
| 21 | 49 | 67.1 | 645  | 3  | P91385 | 2.53e+01 |
| 22 | 49 | 67.1 | 1061 | 4  | P77591 | 2.53e+01 |
| 23 | 49 | 67.1 | 1199 | 4  | Q28205 | 2.53e+01 |
| 24 | 48 | 65.8 | 98   | 11 | 011494 | 3.92e+01 |
| 25 | 48 | 65.8 | 156  | 9  | 058887 | 3.92e+01 |
| 26 | 48 | 65.8 | 219  | 9  | P74029 | 3.92e+01 |
| 27 | 48 | 65.8 | 293  | 9  | 059603 | 3.92e+01 |
| 28 | 48 | 65.8 | 311  | 9  | Q48045 | 3.92e+01 |
| 29 | 48 | 65.8 | 318  | 9  | Q30868 | 3.92e+01 |
| 30 | 48 | 65.8 | 663  | 3  | 016057 | 3.92e+01 |
| 31 | 48 | 65.8 | 692  | 9  | 034448 | 3.92e+01 |
| 32 | 48 | 65.8 | 1381 | 11 | 066628 | 3.92e+01 |
| 33 | 47 | 64.4 | 132  | 11 | 036882 | 6.02e+01 |
| 34 | 47 | 64.4 | 132  | 11 | 036877 | 6.02e+01 |
| 35 | 47 | 64.4 | 193  | 9  | 033174 | 6.02e+01 |
| 36 | 47 | 64.4 | 196  | 9  | 027375 | 6.02e+01 |
| 37 | 47 | 64.4 | 319  | 2  | 014626 | 6.02e+01 |
| 38 | 47 | 64.4 | 329  | 11 | 065253 | 6.02e+01 |
| 39 | 47 | 64.4 | 342  | 9  | P71781 | 6.02e+01 |
| 40 | 47 | 64.4 | 364  | 9  | P73423 | 6.02e+01 |
| 41 | 47 | 64.4 | 568  | 3  | Q27212 | 6.02e+01 |
| 42 | 47 | 64.4 | 662  | 3  | P78695 | 6.02e+01 |
| 43 | 47 | 64.4 | 672  | 1  | 014453 | 6.02e+01 |
| 44 | 47 | 64.4 | 672  | 1  | 013280 | 6.02e+01 |
| 45 | 47 | 64.4 | 698  | 9  | Q55736 | 6.02e+01 |

## ALIGNMENTS

| RESULT   | ID   | PRELIMINARY | PRT | 410 AA. |
|--|--|-------------|-----|---------|
| 1  | 033331   |             |     |         |
| AC   | 033331   |             |     |         |
| DT   | 01-JAN-1998 (TREMBLREL. 05, CREATED)                               |             |     |         |
| DT   | 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)                  |             |     |         |
| DT   | 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)                |             |     |         |
| DE   | DEHYDROGENASE.   |             |     |         |
| GN   | MYV002.54C   |             |     |         |
| OS   | MYCOBACTERIUM TUBERCULOSIS.  |             |     |         |
| OC   | PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.         |             |     |         |
| RP   | SEQUENCE FROM N.A.   |             |     |         |
| RP   | STRAIN-H37RV.  |             |     |         |
| RA   | MDPRY L., HARRIS D.,   |             |     |         |
| RL   | SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.               |             |     |         |
| RL   | [2]  |             |     |         |
| RP   | SEQUENCE FROM N.A.   |             |     |         |
| RC   | STRAIN-H37RV.  |             |     |         |
| RA   | PARKHILL J., BARRELL B.G., RAJANDREAM M.A.,                        |             |     |         |
| RL   | SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.               |             |     |         |
| RL   | [3]  |             |     |         |
| RP   | SEQUENCE FROM N.A.   |             |     |         |
| RC   | STRAIN-H37RV.  |             |     |         |
| RX   | MEDLINE; 96181548.   |             |     |         |
| RA   | PHILIPP W.J., POULET S., EIGMEIER K., PASCOPELLA L.,               |             |     |         |
| RA   | BALASBRAMANIAN V., HEYM B., BERSH S., BLOOM B.R., JACOBS W.R. JR., |             |     |         |
| RA   | COLE S.T.,   |             |     |         |
| RL   | PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).                  |             |     |         |
| DR   | EMBL; AL008967; E1173919; -  |             |     |         |
| SO   | SEQUENCE 410 AA; 44743 MW; 9BA843CB CIRC32;                        |             |     |         |
| Query Match  |  |             |     |         |
| Best Local Similarity 80.8%; Score 59; DB 9; Length 410;   |  |             |     |         |
| Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0; |  |             |     |         |
| Db   | 37 PYPIARL 44  |             |     |         |
| QY   | 2 PYPIARL 9  |             |     |         |
| RESULT   | 2  | PRELIMINARY | PRT | 52 AA.  |
| ID   | 032283   |             |     |         |
| AC   | 032283   |             |     |         |
| DT   | 01-JAN-1998 (TREMBLREL. 05, CREATED)                               |             |     |         |

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE YXZF PROTEIN.  
 GN YXZF  
 OS BACILLUS SUBTILIS.  
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168:  
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,  
 AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,  
 BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,  
 BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,  
 CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,  
 DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,  
 ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOUTER D., FRITZ C.,  
 FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLEON N., GHIM S.Y.,  
 GLASER P., GOFEAU A., GOLIGHTLY E.J., GRANDI G., GUISSEPI G., GUY B.J.,  
 HAGA K., HALECH J., HARWOOD C.R., HENAUT A., HILBERT H., HOLSAPPEL S.,  
 HOSONO S., HUILO M.F., ITAYA M., JONES L., JORIS B., KARAWATA D.,  
 KASABARA Y., KLAER-BLANCHARD M., KLEIN C., KOBAYASHI Y., KOETTER P.,  
 KONIGSSTEIN G., KROGH S., KOMANO M., KURITA K., LAPIDUS A.,  
 LARDINOIS S., LAUBER J., LAZAREVIC V., LEE S.M., LEVINE A., LIU H.,  
 MASUDA S., MAUL C., MEDIGUE C., MEDINA N., MELLADO R.P., MIZUNO M.,  
 MOESTL D., NAKAI S., NOBACK M., NOONE D., O'REILLY M., OGAWA K.,  
 OGIMARA A., OUDEGA B., PARK S.H., PARO V., POHL T.M., PORTETELLE D.,  
 POROLITK S., PRESCOTT A.M., PRESECAN E., PUTIC P., PURNELLE B.,  
 RAPPORT G., REY M., REYNOLDS S., RIEGER M., RIVOLTA C., ROCHA E.,  
 ROCHE B., ROSE M., SADAIE Y., SATO T., SCANLAN E., SCHLEICH S.,  
 SCHROETER R., SCOFONE F., SEKIGUCHI J., SEKOSKA A., SERO S.J.,  
 SERROR P., SHIN B.S., SORDO B., SOROKIN A., TACCONE I., TAKAGI T.,  
 TAKAHASHI H., TAKEKAWA K., TAKEUCHI M., TAMAKOSHI A., TANAKA T.,  
 TERPETA P., TOGNONI A., TOSATO V., UCHIYAMA S., VANDEBOL M.,  
 VANNIER F., VASSAROTTI A., VIARI A., WAMBITT R., WEDLER E., WEDLER H.,  
 WEITZENGGER T., WINTERS P., WIPAT A., YAMAMOTO H., YANANE K.,  
 YASHIKAWA H., YATA K., YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E.,  
 RA YOSHAKAWA H., DANCHIN A.;  
 RL NATURE 390:249-256(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168:  
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: 299123; E1186360;  
 SO SEQUENCE 52 AA; 5915 MW; 8AE33D5F CRC32;

Query Match 72.6%; Score 53; DB 9; Length 52;  
 Best Local Similarity 66.7%; Pred. No. 4.13e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 25 VYPIVRI 33  
 1 IYPIVRI 9

RESULT 3  
 ID P87115 PRELIMINARY; PRT; 1033 AA.  
 AC P87115;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 116.5 KD PROTEIN.  
 GN SPAC2068.09C  
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCOMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972H-;  
 RA BADCOCK K., CHURCHER C.M.;  
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972H-;

RA WOOD V., BARRELL B.G., RAJANDREAM M.A.;  
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: 295334; E315494;  
 KW HYPOTHETICAL PROTEIN.  
 SO SEQUENCE 1033 AA; 116463 MW; ACDAC931 CRC32;

Query Match 72.6%; Score 53; DB 1; Length 1033;  
 Best Local Similarity 77.8%; Pred. No. 4.13e+00;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 391 IPLYVRI 399  
 1 IPLYVRI 9

RESULT 4  
 ID P74054 PRELIMINARY; PRT; 451 AA.  
 AC P74054;  
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
 DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 50.4 KD PROTEIN.  
 OS SYNECHOCYSTIS SP.  
 OC EUBACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RA TABATA S.;  
 RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
 MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAKOTO S., KIMURA T.,  
 RA HOSOUCHI T., MATSUNO A., MORIKI A., NAKAZAKI N., NARO K.,  
 RA OKUMURA S., SHIMO S., TAKEUCHI C., WADA T., WATANABE A.,  
 RA YAMADA M., YASUDA M., TABATA S.;  
 RL DNA RES. 3:109-136(1996).  
 DR EMBL: D90911; G1553214;  
 KW HYPOTHETICAL PROTEIN.  
 SO SEQUENCE 451 AA; 50417 MW; 42DCE091 CRC32;

Query Match 71.2%; Score 52; DB 9; Length 451;  
 Best Local Similarity 62.5%; Pred. No. 6.56e+00;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 110 LPYMYR 117  
 1 IYPIVRI 8

RESULT 5  
 ID O31936 PRELIMINARY; PRT; 75 AA.  
 AC O31936;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE YOBP PROTEIN.  
 GN YOBP.  
 OS BACILLUS SUBTILIS.  
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,  
 AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,  
 BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,  
 BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,  
 CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,  
 DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,  
 ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOUTER D., FRITZ C.,  
 FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLEON N., GHIM S.Y.,  
 GLASER P., GOFEAU A., GOLIGHTLY E.J., GRANDI G., GUISSEPI G., GUY B.J.,



RA HAGA K., HALECH J., HARWOOD C.R., HENAUT A., HILBERT H., HOLSAPPEL S.,  
 RA HOSONO S., HULLO M.F., ITAYA M., JONES L., JORIS B., KARAMATA D.,  
 RA KASAHARA Y., KLAER-BLANCHARD M., KLEIN C., KOBAYASHI Y., KOETTER P.,  
 RA KINGSTON G., KROCH S., KUWANO M., KURITA K., LAPIDUS A.,  
 RA LARDINOIS S., LAUBER J., LAZAREVIC V., LEE S.M., LEVINE A., LIU H.,  
 RA MASUDA S., MAUEL C., MEDIGUE C., MEDINA N., MELLAO R.P., MIZUNO M.,  
 RA MESTL D., NAKEL S., NOBACK M., NOONE D., O'REILLY M., OGAWA K.,  
 RA OCHIMAYA A., OUDGA B., PARK S.H., PARRO V., POHL T.M., PORTELELLA D.,  
 RA POROOLIK S., PRESCOTT A.M., PRESECAN E., PUTIC P., PUNNELLE B.,  
 RA RAPPORT G., REY M., REYNOLDS S., RIEGER M., RIYOLIA C., ROCHA E.,  
 RA ROCHE B., ROSE M., SADAIE Y., SATO T., SCANLAN E., SCHLEICH S.,  
 RA SCHROETER R., SCOPFONE F., SEKIGUCHI J., SEKOWSKA A., SERO S.J.,  
 RA SERROR P., SHIN B.S., SOLDO B., SOROKIN A., TACCONI E., TAKAGI T.,  
 RA TAKAHASHI H., TAKEWAKU K., TAKEUCHI M., TAMAKOSHI A., TANAKA T.,  
 RA TERPSTRA P., TOGNONI A., TOSATO V., UCHIYAMA S., VANDERBOL M.,  
 RA VANIER F., VASAROTTI A., VIARI A., VAMBOTI R., WEDLER E., WEDLER H.,  
 RA WEITZENEGGER T., WINTERS P., WIPPT A., YAMAMOTO H., YAMANE K.,  
 RA YASUMOTO K., YATA K., YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E.,  
 RA YOSHIKAWA H., DANCHIN A.,  
 RL NATURE 390:249-256(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.,  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: 299115; E1183542;  
 SQ SEQUENCE 75 AA; 9099 MW; 115189B3 CRC32;  
  
 Query Match 69.9%; Score 51; DB 9; Length 75;  
 Best Local Similarity 85.7%; Pred. No. 1.04e+01;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
 Db 28 YPTVYRKL 34  
 QY 3 YPTVYRKL 9  
  
 RESULT 6  
 ID 084619; PRELIMINARY; PRT; 112 AA.  
 AC 084619;  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DE GENOME, PARTIAL SEQUENCE.  
 GN A3033.  
 OS PARAECTIUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).  
 OC VIRIDAE: DS-DNA NONENVELOPED VIRUSES; PHYCODNAVIRIDAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95133167.  
 RA LU Z., LI Y., ZHANG Y., KUTISH G.F., ROCK D.L., VAN ETTEN J.L.,  
 RL VIROLOGY 206:339-352(1995).  
 DR EMBL: U42580; G1181466;  
 SQ SEQUENCE 112 AA; 13416 MW; 5C07006C CRC32;  
  
 Query Match 69.9%; Score 51; DB 11; Length 112;  
 Best Local Similarity 44.4%; Pred. No. 1.04e+01;  
 Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
  
 Db 20 VPSIIRNL 28  
 QY 1 IYPIYRKL 9  
  
 RESULT 7  
 ID 009006; PRELIMINARY; PRT; 133 AA.  
 AC 009006;  
 DT 01-JUL-1997 (TREMELREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMELREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)  
 DE SMALL INDUCIBLE CYTOKINE A21 (BETA CHEMOKINE EXODUS-2).  
 GN SCYA21.  
 OS MUS MUSCULUS (MOUSE).

OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA;  
 OC EUTHERIA: RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TOTAL FETUS;  
 RA HROMAS R.A., GRAY P., KLEMSZ M., FIFE K., BROXMEYER H.,  
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U88322; G2196924;  
 MGD: MGI:1097677; SCYA21.  
 SQ SEQUENCE 133 AA; 14615 MW; FEFASB6 CRC32;  
  
 Query Match 69.9%; Score 51; DB 10; Length 133;  
 Best Local Similarity 85.7%; Pred. No. 1.04e+01;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
 Db 40 IYPIYR 46  
 QY 1 IYPIYR 7  
  
 RESULT 8  
 ID 009002; PRELIMINARY; PRT; 133 AA.  
 AC 009002;  
 DT 01-JUL-1997 (TREMELREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMELREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)  
 DE TCM.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA;  
 OC EUTHERIA: RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-THIMOS;  
 RA TANABE S., LU Z., LIO Y., QUACKENBUSH E.J., BERMAN M.A.,  
 RA COLLINS-RACIE L.A., MI S., REILLY C., LO D., JACOBS K.A., DORE M.E.,  
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 97400322.  
 RA HEDRICK J.A., ZLOTNIK A.,  
 RL J. IMMUNOL. 159:1589-1593(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA HEDRICK J.A., ZLOTNIK A.,  
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AF006637; G2209189;  
 DR EMBL: AF001980; G2624927;  
 SQ SEQUENCE 133 AA; 14558 MW; C0532523 CRC32;  
  
 Query Match 69.9%; Score 51; DB 10; Length 133;  
 Best Local Similarity 85.7%; Pred. No. 1.04e+01;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
 Db 40 IYPIYR 46  
 QY 1 IYPIYR 7  
  
 RESULT 9  
 ID 058389; PRELIMINARY; PRT; 197 AA.  
 AC 058389;  
 DT 01-JAN-1998 (TREMELREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)  
 DE HYPOTHETICAL PROTEIN MJ0979.  
 GN MJ0979.  
 OS METHANOCOCCUS JANNASCHII.  
 OC ARCHAEABACTERIA: EURYARCHAEOTA: METHANOCOCCALES: METHANOCOCCACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96337999.  
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,

RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
 RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,  
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KATNE B.P., BOBODOVSKI M.,  
 RA KLEINK H.-P., FRASER C.M., SMITH H.O., WOESSE C.R., VENTER J.C.,  
 RA SCIENCE 273:1058-1073(1996).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 DR EMBL: U67541; G1499818; -  
 KM HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
 FT TRANSMEM 11 31 POTENTIAL.  
 FT TRANSMEM 85 105 POTENTIAL.  
 FT TRANSMEM 109 129 POTENTIAL.  
 FT TRANSMEM 174 194 POTENTIAL.  
 SO SEQUENCE 197 AA; 21520 MW; 926ECAA9 CRC32;

Query Match 69.9%; Score 51; DB 9; Length 197;  
 Best Local Similarity 55.6%; Pred. No. 1.04e+01;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 189 IAYPIRKV 197  
 QY 1 IPIPIVKRL 9

RESULT 10  
 ID 051549 PRELIMINARY; PRT: 766 AA.  
 AC 051549;  
 DT 01-JAN-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE KILLER PROTEIN OF PYOCIN S3.  
 GN PYOSA.  
 OS PSEUDOMONAS AERUGINOSA.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
 CC PSEUDOMONADACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-P12;  
 RX MEDLINE: 95238389.  
 RA DUPOURT C., BAYSSE C., MICHEL-BRIAND Y.,  
 RA J. BIOL. CHEM. 270:8920-8927(1995).  
 DR EMBL: X77996; G854363; -  
 SO SEQUENCE 766 AA; 81434 MW; 66F2A86E CRC32;

Query Match 69.9%; Score 51; DB 9; Length 766;  
 Best Local Similarity 66.7%; Pred. No. 1.04e+01;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 679 IYGEIRKL 687  
 QY 1 IPIPIVKRL 9

RESULT 11  
 ID 036900 PRELIMINARY; PRT: 129 AA.  
 AC 036900;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE MA-P17 (FRAGMENT).  
 GN GAG.  
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 CC LENTIVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA LEIGH BROWN A.J., LOBIDEL D., MADE C.M., REBUS S., PHILLIPS N.,  
 RA BRETTLE R.P., FRANCE A.J., LEEN C.S., MCENAMIN J., MCWILLAN A.,  
 RA MAW R.D., MURCAHY F., ROBERTSON J.R., SANKAR K.N., SCOTT G., WYLD R.,  
 RA PEUTHERER J.F.,  
 RA VIROLOGY 235:166-177(1997).  
 DR EMBL: AF014297; G2406936; -

FT NON\_TER 1 1  
 FT NON\_TER 129  
 SO SEQUENCE 129 AA; 14436 MW; 5201F732 CRC32;

Query Match 68.5%; Score 50; DB 11; Length 129;  
 Best Local Similarity 85.7%; Pred. No. 1.63e+01;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 109 YPIVOKL 115  
 QY 3 YPIVOKL 9

RESULT 12  
 ID 036786 PRELIMINARY; PRT: 130 AA.  
 AC 036786;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE MA-P17 (FRAGMENT).  
 GN GAG.  
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 CC LENTIVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA LEIGH BROWN A.J., LOBIDEL D., MADE C.M., REBUS S., PHILLIPS N.,  
 RA BRETTLE R.P., FRANCE A.J., LEEN C.S., MCENAMIN J., MCWILLAN A.,  
 RA MAW R.D., MURCAHY F., ROBERTSON J.R., SANKAR K.N., SCOTT G., WYLD R.,  
 RA PEUTHERER J.F.,  
 RA VIROLOGY 235:166-177(1997).  
 DR EMBL: AF014183; G2406708; -  
 FT NON\_TER 1 1  
 FT NON\_TER 130  
 SO SEQUENCE 130 AA; 14584 MW; 428D9E75 CRC32;

Query Match 68.5%; Score 50; DB 11; Length 130;  
 Best Local Similarity 85.7%; Pred. No. 1.63e+01;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 109 YPIVOKL 115  
 QY 3 YPIVOKL 9

RESULT 13  
 ID 020408 PRELIMINARY; PRT: 280 AA.  
 AC 020408;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE F44F4.10.  
 OS CAENORHABDITIS ELEGANS.  
 CC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CODES L.,  
 RA SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BEKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J.,  
 RA COULSON A., CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,  
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
 RA LARREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORITMORE B.,  
 RA O'CALLAGHAN M., PARSONS J., PERCY C., RITKEN L., ROOPRA A.,  
 RA SANDERS D., SHOWNKEEN R., SVALDON N., SMITH A., SONNHAMMER E.,  
 RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
 RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
 RA WILKINSON-SPROAT J., WOHLDMAN P.,  
 RA NATURE 368:32-38(1994).

DR EMBL: Z37092; G558287;  
SQ SEQUENCE 280 AA; 31710 MW; F98B0CB8 CRC32;

Query Match 68.5%; Score 50; DB 3; Length 280;  
Best Local Similarity 100.0%; Pred. No. 1.63e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 241 YPIVRK 246  
|||  
QY 3 YPIVRK 8

RESULT 14  
ID 006236 PRELIMINARY; PRT; 357 AA.

AC 006236;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE PYRD.  
GN. PYRD.

OS MYCOBACTERIUM TUBERCULOSIS.  
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA BARRELL K., CHURCHER C.M.;  
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA BARRELL B.G., RAJANDREAM M.A., PARKHILL J.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE: 96181548.

RA PHILLIP W.J., POULET S., EIGMEYER K., PASCOPELLA L.,  
RA BALASUBRAMANIAN V., HEYM B., BENGH S., BLOOM B.R., JACOBS W.R. JR.,  
RA COLE S.T.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).  
DR EMBL: Z95388; E316034;  
DR PROSITE: PS00911; RHODEHASE\_1; 1.  
DR PROSITE: PS00912; RHODEHASE\_2; 1.  
SQ SEQUENCE 357 AA; 37998 MW; 56358C06 CRC32;

Query Match 68.5%; Score 50; DB 9; Length 357;  
Best Local Similarity 71.4%; Pred. No. 1.63e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 2 YPIVRK 8  
|||  
QY 3 YPIVRK 9

RESULT 15  
ID 079898 PRELIMINARY; PRT; 491 AA.

AC 079898;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE GAG PROTEIN.

GN GAG.  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
OC LENTIVIRINAE.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PH136; TISSUE-BLOOD;  
RA LOUWAGIE J.J., MCCUTCHAN F., BRENNAN T., PEETERS M., BRENNAN T.,  
RA SANDERS-BUELL E., EDDY G., DER GROEN G., FRANSSEN K.,  
RA GERSHAY-DAMET M., DELEYS R., BURKE D.;  
RL SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: L11780; G554991;

SQ SEQUENCE 491 AA; 54891 MW; 657C69DA CRC32;

Query Match 68.5%; Score 50; DB 11; Length 491;  
Best Local Similarity 85.7%; Pred. No. 1.63e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 132 YPIVRK 138  
|||  
QY 3 YPIVRK 9

Search completed: Fri Sep 11 13:07:52 1998  
Job time : 32 secs.

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
 W E S E R E H  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPerch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 13:01:42 1998; MasPar time 2.75 Seconds  
 53.037 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-7  
 Description: (1-9) from US08452843.pep  
 Perfect Score: 65  
 Sequence: 1 QPDDAVYKL 9

Scoring table:  
 PAM 150  
 Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a:geneseq32  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29

Statistics: Mean 16.624; Variance 46.989; scale 0.354

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID     | Description            | Pred. No. |
|------------|-------|-------------|--------|--------|------------------------|-----------|
| 1          | 65    | 100.0       | 9      | R89368 | Cw4 consensus peptide  | 7.69e-01  |
| 2          | 50    | 76.9        | 771    | R71280 | Human semaphorin III   | 4.63e-01  |
| 3          | 49    | 75.4        | 584    | R14670 | Truncated poly Ig-rec  | 6.02e-01  |
| 4          | 49    | 75.4        | 773    | W03177 | Rabbit poly-immunoglob | 6.02e-01  |
| 5          | 45    | 69.2        | 198    | R05881 | Sequence encoded by c  | 1.68e-02  |
| 6          | 45    | 69.2        | 580    | R38151 | Acetobacter diquanyla  | 1.68e-02  |
| 7          | 44    | 67.7        | 653    | R06723 | Acetobacter proteas    | 2.17e-02  |
| 8          | 43    | 66.2        | 108    | R33271 | T cell receptor alpha  | 2.78e-02  |
| 9          | 43    | 66.2        | 112    | R33270 | Human T-cell receptor  | 2.78e-02  |
| 10         | 43    | 66.2        | 112    | R33270 | Human T-cell receptor  | 2.78e-02  |
| 11         | 43    | 66.2        | 113    | W36108 | Human T-cell receptor  | 2.78e-02  |
| 12         | 43    | 66.2        | 114    | R33269 | Human T-cell receptor  | 2.78e-02  |
| 13         | 43    | 66.2        | 114    | R33276 | T cell receptor alpha  | 2.78e-02  |
| 14         | 43    | 66.2        | 117    | R33275 | T cell receptor alpha  | 2.78e-02  |
| 15         | 43    | 66.2        | 117    | R33275 | T cell receptor alpha  | 2.78e-02  |
| 16         | 43    | 66.2        | 117    | R33274 | T cell receptor alpha  | 2.78e-02  |
| 17         | 43    | 66.2        | 117    | R33272 | T cell receptor alpha  | 2.78e-02  |
| 18         | 43    | 66.2        | 136    | R33277 | T cell receptor alpha  | 2.78e-02  |

| Result | Score | Query Match | Length | ID     | Description           | Pred. No. |
|--------|-------|-------------|--------|--------|-----------------------|-----------|
| 19     | 43    | 66.2        | 277    | P60065 | Sequence of a polypep | 2.78e-02  |
| 20     | 43    | 66.2        | 306    | R12346 | Toxoplasma gondii pro | 2.78e-02  |
| 21     | 43    | 66.2        | 306    | W01731 | T. gondii antigen par | 2.78e-02  |
| 22     | 43    | 66.2        | 422    | W01732 | T. gondii antigen p68 | 2.78e-02  |
| 23     | 43    | 66.2        | 452    | R12353 | Toxoplasma gondii p68 | 2.78e-02  |
| 24     | 43    | 66.2        | 593    | R28349 | Bacillus caldodenax D | 2.78e-02  |
| 25     | 43    | 66.2        | 876    | W35905 | Bacillus steatochemo  | 2.78e-02  |
| 26     | 43    | 66.2        | 877    | W22846 | Bacillus caldodenax D | 2.78e-02  |
| 27     | 43    | 66.2        | 877    | W22845 | Bacillus caldodenax D | 2.78e-02  |
| 28     | 43    | 66.2        | 877    | R28348 | Bacillus caldodenax D | 2.78e-02  |
| 29     | 43    | 66.2        | 877    | W22847 | Bacillus caldodenax D | 2.78e-02  |
| 30     | 43    | 66.2        | 1074   | R24102 | Marek's disease virus | 2.78e-02  |
| 31     | 43    | 66.2        | 2386   | W13153 | S. pombe Rad3 polyep  | 2.78e-02  |
| 32     | 43    | 66.2        | 109    | W36107 | Mouse T-cell receptor | 3.57e-02  |
| 33     | 42    | 64.6        | 215    | R77287 | T-cell receptor alpha | 3.57e-02  |
| 34     | 42    | 64.6        | 267    | W04300 | Murine T-cell recepto | 3.57e-02  |
| 35     | 42    | 64.6        | 268    | W36110 | Mouse T-cell receptor | 3.57e-02  |
| 36     | 42    | 64.6        | 272    | W36111 | Mouse T-cell receptor | 3.57e-02  |
| 37     | 42    | 64.6        | 464    | R50197 | Human wild type gluc  | 3.57e-02  |
| 38     | 42    | 64.6        | 465    | R50195 | Human wild type gluc  | 3.57e-02  |
| 39     | 42    | 64.6        | 465    | W37432 | Rat liver glucokinase | 3.57e-02  |
| 40     | 42    | 64.6        | 465    | W37438 | Rat liver glucokinase | 3.57e-02  |
| 41     | 42    | 64.6        | 521    | W34477 | Rat islet glucokinase | 3.57e-02  |
| 42     | 42    | 64.6        | 855    | R75505 | RCH1-related proteina | 3.57e-02  |
| 43     | 42    | 64.6        | 855    | R75503 | Human colonic adenoca | 3.57e-02  |
| 44     | 42    | 64.6        | 919    | W37436 | Hexokinase-glucokinas | 3.57e-02  |
| 45     | 42    | 64.6        | 919    | W37434 | Hexokinase-glucokinas | 3.57e-02  |

## ALIGNMENTS

RESULT 1  
 ID R89368 standard; peptide: 9 AA.  
 AC R89368;  
 DT 18-SEP-1996 (first entry)  
 DE Cw4 consensus peptide derived immunogenic peptide.  
 KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
 KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
 KW hepatitis C;  
 OS Synthetic.  
 PN W09603140-A1.  
 PD 08-FEB-1996.  
 PF 21-JUL-1995; U09234.  
 PR 21-JUL-1994; US-278634.  
 PR 23-NOV-1994; US-344824.  
 PR 30-MAY-1995; US-452843.  
 PA (CYTE-) CYTEL CORP.  
 PI Sette A, Sidney J;  
 DR WPI: 96-116784/12.  
 PT Compsn. comprising immunogenic peptide with supermotif allowing more  
 PT than one HLA mol. to bind - used to induce CTL response in patient  
 PT and for in vivo and ex vivo therapeutic and diagnostic applications  
 PS Claim 2: Page 26; 32pp; English.  
 CC The sequences given in R89362-82 are immunogenic peptides which were  
 CC use in the composition of the invention. The composition comprises  
 CC an immunogenic peptide of 9-10 residues with a supermotif which  
 CC allows binding of more than one HLA molecule. It pref. comprises  
 CC two conserved residues, a first at the 2nd position from the N-  
 CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
 CC are used to induce a CTL response in a patient. They are also  
 CC useful in compositions for in vivo and ex vivo therapeutic and  
 CC diagnostic applications, e.g. hepatitis B and C.  
 CC infections, e.g. hepatitis B and C.  
 SQ Sequence 9 AA;

Query Match 100.0%; Score 65; DB 18; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 7.69e-01;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 1 qpddavykl 9  
 |||||  
 1 qpddavykl 9

RESULT 2  
 ID R71380 standard; Protein: 771 AA.  
 AC R71380;  
 DT 21-NOV-1995 (first entry)  
 DE Human semaphorin III protein.  
 KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;  
 KW varicella major virus; smallpox; semaphorin receptor binding activity;  
 KW modulation; nerve cell growth; immune response; viral pathogenesis;  
 KW neurological disease; neuro-regeneration; oncological infection.  
 OS Homo sapiens.  
 PN WO9507706-A.  
 PD 23-MAR-1995.  
 PE 13-SEP-1994; U10151.  
 PR 13-SEP-1993; US-121713.  
 RA (RGC) UNIV CALIFORNIA.  
 PI Beptley DR, Goodman CS, Kolodkin AL, Matthes D;  
 PI O'Connor T;  
 PI WPI: 95-131177/17.  
 DR N-PSDB: 087442.  
 PT New class of semaphorin peptide(s) and polypeptide(s) - are  
 PT potent modulators of nerve cell growth and regeneration  
 PS Example 2: Page 60-63; 101pp; English.  
 CC The sequence of the human semaphorin III protein. The proteins  
 CC encoded by the grasshopper semaphorin I (087441), human semaphorin III,  
 CC vaccinia virus semaphorin IV (087443), Drosophila semaphorin I and II  
 CC (087444-5), Tribolium semaphorin I (087446) or varicella major (smallpox)  
 CC virus semaphorin IV (087447) genes were used to generate a series of  
 CC peptides (R70370-R70418), which retain semaphorin receptor binding  
 CC activity. The semaphorin derived or semaphorin receptor derived peptides  
 CC are potent modulators of nerve cell growth, immune responsiveness and  
 CC viral pathogenesis. They can be used in diagnosis and treatment of  
 CC neurological disease and neuro-regeneration, immune modulation and  
 CC diagnosis and treatment of viral and oncological infection and diseases.  
 CC Sequence 771 AA:  
 SO

Query Match 76.9%; Score 50; DB 13; Length 771;  
 Best Local Similarity 44.4%; Pred. No. 4.63e+01;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 149 hpednifk1 157  
 :|:|:|:  
 OY 1 QPDDAVYKL 9

RESULT 3  
 ID R14670 standard; Protein: 584 AA.  
 AC R14670;  
 DT 30-JAN-1992 (first entry)  
 DE Truncated poly Ig-receptor encoded by allele no. 1.  
 KW Rabbit; insemination; pregnancy.  
 OS Oryctolagus cuniculus.  
 FH Key  
 FT peptide  
 FT 1:18  
 FT Location/Qualifiers  
 FT /label= signal sequence  
 FT 10..118  
 FT /number= I  
 FT 119..223  
 FT /note= "Poly-Ig binding"  
 FT /number= II  
 FT 224..332  
 FT /number= III  
 FT 333..441  
 FT /number= IV  
 FT 442..552  
 FT /number= V  
 FT 553..584  
 FT /number= IV  
 FT /note= "incomplete"  
 PN WO9116061-A.  
 PD 31-OCT-1991.  
 PE 16-APR-1991; U02604.  
 PR 16-APR-1990; US-510161.

PA (HARD) HARVARD COLLEGE.  
 PA (SURE-) INST SUISSE RECH. EXPR C.  
 PI Kriehenbuhl JP, Weltzin RA, Neutra MR;  
 DR WPI: 91-339549/46.  
 DR N-PSDB: Q14498.  
 PT Stabilised poly-Ig complex contg. portion of poly-Ig receptor -  
 PT useful in protection against pathogens or against pregnancy  
 PS Disclosure: Fig 7: 51 pp; English.  
 CC The sequence was deduced from a cDNA clone of allele no. 1 and  
 CC is a truncated poly-Ig receptor. The native gene (Mostov et al)  
 CC is mutated to delete the portion encoding the transmembrane and  
 CC intra-cellular domains. The recombinant protein produced by  
 CC expression of the sequence is used as a stabiliser protein with a  
 CC poly-Ig specific for a selected antigen or family of antigens. The  
 CC compsn. can be administered directly to the mucosal surfaces of a  
 CC mammal to protect against a pathogen or against insemination. It  
 CC protects against allergens that contact the respiratory or digestive  
 CC mucosal surfaces and protects against pregnancy by cross-linking  
 CC sperm in the vagina.  
 CC See also R14671.  
 CC Sequence 584 AA;  
 SO

Query Match 75.4%; Score 49; DB 3; Length 584;  
 Best Local Similarity 75.0%; Pred. No. 6.02e+01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 137 epdvvvk 144  
 :|:|:|:  
 OY 1 QPDDAVYK 8

RESULT 4  
 ID W03177 standard; Protein: 773 AA.  
 AC W03177;  
 DT 24-FEB-1997 (first entry)  
 DE Rabbit poly-immunoglobulin receptor.  
 KW Rabbit; immunoglobulin; receptor; protection protein; mutants;  
 KW heavy chain; antigen binding domain; protection; pathogen;  
 KW mucosal; environment; gastrointestinal; passive; immunisation;  
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;  
 KW poly; sordinus.  
 OS Oryctolagus cuniculus.  
 FH Key  
 FT region  
 FT 21..43  
 FT Location/Qualifiers  
 FT /note= "Immunoglobulin binding residues of  
 FT domain I"  
 FT 1..118  
 FT /label= domain\_I  
 FT 119..223  
 FT /label= domain\_II  
 FT 224..332  
 FT /label= domain\_III  
 FT 333..441  
 FT /label= domain\_IV  
 FT 442..552  
 FT /label= domain\_V  
 FT 553..606  
 FT /note= "external portions of domain VI"  
 FT 553..627  
 FT /note= "external portions of domain VI"  
 FT 630..652  
 FT /label= transmembrane-segment  
 FT 653..755  
 FT /label= intracellular\_portion  
 FT region  
 PN WO9621012-A1.  
 PD 11-JUL-1996.  
 PE 27-DEC-1995; U16889.  
 PR 30-DEC-1994; US-367395.  
 PR 04-MAY-1995; US-434000.  
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.  
 PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
 PA (PLAN-) PLANET BIOTECHNOLOGY INC.  
 PI Hlatk AC, Lehner T, Ma JKC.

DR MPI: 96-333987/33.  
 DR N-PSDB: T31287.  
 PT Immunoglobulin and protection protein complex and its prodn. in  
 plants - useful for passive immunisation against mucosal antigens,  
 esp. against S. mutans and S. sorbinus to prevent dental caries  
 PS Claim 10; Pages 99-102; 152pp; English.  
 CC The present sequence is the rabbit poly-immunoglobulin (Ig)  
 receptor, a portion of which corresp. to residues 1-627, pref.  
 CC 1-606, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,  
 CC 442-552, 553-606 or 553-627 comprises a protection protein (PP).  
 CC The Ig of the invention comprises a PP as above in association with  
 CC an Ig derived heavy chain, having at least a portion of an antigen  
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.  
 CC gastrointestinal, environments, therefore enhancing its  
 CC effectiveness in passively immunising animals against mucosal  
 CC pathogens. The Ag binding domain is specifically derived from the  
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries  
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or  
 CC S. sorbinus serotypes d and g.  
 SQ Sequence 773 AA;  
 Query Match 75.4%; Score 49; DB 20; Length 773;  
 Best Local Similarity 75.0%; Pred. No. 6.02e+01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Db 137 epdavyk 144  
 QY 1 QPDDAVYK 8  
 RESULT  
 ID R05881 standard; protein: 198 AA.  
 AC R05881;  
 DT 27-NOV-1990 (first entry)  
 DE Sequence encoded by clone 61.  
 KW Interleukin-2; IL-2; cancer; ds.  
 OS Homo sapiens.  
 PN US4939093-A.  
 PD 03-JUL-1990.  
 PF 23-AUG-1988; 236296.  
 PR 28-SEP-1982; US-426059.  
 PR 13-JAN-1983; US-457594.  
 PR 02-FEB-1987; US-009999.  
 PR 23-AUG-1988; US-236296.  
 PA (CERU) CERUS CORP.  
 PI MCGROGAN MP, KAWASAKI ES, DOYLE MV, MARK DF;  
 DR N-PSDB: 005237.  
 DR MPI: 90-224018/29.  
 PT Messenger RNA expressing interleukin 2 in X.laavis oocyte -  
 PT isolated by hybridisation with new recombinant DNA, also useful  
 PT for expression in bacterial hosts.  
 PS Disclosure; P; English.  
 CC Clone may be used to produce IL-2 in a X.laavis oocyte translation  
 CC system. IL-2 is useful in diagnosis and treatment of cancer,  
 CC infections and immune diseases.  
 SQ Sequence 198 AA;  
 Query Match 69.2%; Score 45; DB 1; Length 198;  
 Best Local Similarity 62.5%; Pred. No. 1.68e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 151 peekykl 58  
 QY 2 PDDAVYKL 9  
 RESULT  
 ID R38151 standard; protein: 580 AA.  
 AC R38151;  
 DT 13-OCT-1993 (first entry)  
 DE Acetobacter diacylate cyclase Dgcl.  
 KW Cyclic diacylate; diacylate production; cdgl operon.  
 KW diacylate cyclase; cellulose production; cdgl operon.

OS Acetobacter xylinum.  
 FH Key Location/Qualifiers  
 FT misc.difference 238  
 FT /note= "deduced from CCG (Pro) codon"  
 FT misc.difference 353  
 FT /note= "deduced from CCG (Arg) codon"  
 FT  
 PN W09311244-A.  
 PD 10-JUN-1993.  
 PF 14-OCT-1992; U08756.  
 PR 29-NOV-1991; US-800218.  
 PA (WEYE) WEYERHAEUSER CO.  
 PI Ben-Bassat A, Ben-Ziman M, Calhoun RD, Gelfand DF;  
 PI Tal R, Wong HC;  
 DR MPI: 93-197062/24.  
 DR N-PSDB: Q43660.  
 PT Polynucleotide sequence from Acetobacter cdg operon - encodes  
 PT cyclic di:guanosine mono:phosphate degradation enzymes e.g.  
 PT 3-phosphodiesterase isozyme  
 PS Claim 5; Page 77-79; 98pp; English.  
 CC The amino acid sequence of protein Dgcl was deduced from the third  
 CC open reading frame of the cdgl operon. The protein has diacylate  
 CC cyclase activity, i.e. it enzymatically converts two molecules of  
 CC GTP to bis-(3'5')-cyclic diacylate acid.  
 CC See also R38149-R38150 and R38152.  
 SQ Sequence 580 AA;  
 Query Match 69.2%; Score 44; DB 7; Length 580;  
 Best Local Similarity 55.6%; Pred. No. 1.68e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Db 213 hpedavcrl 221  
 QY 1 QPDDAVYKL 9  
 RESULT  
 ID R06723 standard; protein: 653 AA.  
 AC R06723;  
 DT 18-JAN-1991 (first entry)  
 DE Achromobacter protease I.  
 KW T-API; enzyme prodn.; peptide mapping; peptide synthesis.  
 OS Achromobacter lyticus.  
 PN EP-387646-R.  
 PD 19-SEP-1990.  
 PF 03-MAR-1990; 104163.  
 PF 14-MAR-1989; JP-059726.  
 PA (WAKP) WAKO PURE CHEM IND KK.  
 PI Sakiyama F, Nakata A;  
 DR MPI: 90-283902/38.  
 DR N-PSDB: Q05926.  
 PT Novel DNA encoding Achromobacter protease I - for recombinant  
 PT prodn. of enzyme, and for fragmentation of protein(s) and  
 PT peptide, for peptide mapping and synthesis of lys-x-cpds.  
 PS Disclosure; fig 1; 20pp; English.  
 CC This Achromobacter protease I or an analogue (T-API) specific-  
 CC ally cleaves the peptide bonds (-lys-x-) on the side of the  
 CC carboxyl gps. of lysine residues in proteins and peptides.  
 CC All lys-x bonds are cleaved incl. the lys-pro bond. T-APIs  
 CC are therefore useful for fragmenting proteins or peptides  
 CC for primary structural analysis, prodn. of peptide maps or  
 CC the synthesis of -lys-x- cpds.  
 SQ Sequence 653 AA;  
 Query Match 67.7%; Score 44; DB 1; Length 653;  
 Best Local Similarity 62.5%; Pred. No. 2.17e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 614 pgdvyynl 621  
 QY 2 PDDAVYKL 9  
 RESULT  
 ID R38151 standard; protein: 580 AA.  
 AC R38151;  
 DT 13-OCT-1993 (first entry)  
 DE Acetobacter diacylate cyclase Dgcl.  
 KW Cyclic diacylate; diacylate production; cdgl operon.  
 KW diacylate cyclase; cellulose production; cdgl operon.

ID R33271 standard; Protein: 108 AA.  
 AC R33271:  
 DT 16-JUL-1993 (first entry)  
 DE T cell receptor alpha chain clone alpha1.4.  
 KW Rheumatoid arthritis; synovial; therapy; therapeutic;  
 OS autoimmune response; variable region; mammal; immunisation.  
 PN WO9304695-A.  
 PD 18-MAR-1993.  
 PF 27-AUG-1992; U07289.  
 PR 28-AUG-1991; US-750913.  
 PA (KIRI) UNIV PENNSYLVANIA.  
 PI (WIST-) WISTAR INST.  
 DR Weiner DB, Williams WV;  
 WP1: 93-100655/12.  
 N-PSDB: Q37078.  
 PT T-cell receptor based treatment of rheumatoid arthritis - comprises  
 administration of antibodies to T-cell receptor variable regions  
 PS Disclosure: Page 20; 110pp; English.  
 CC The sequence is that of human rheumatoid synovial T cell receptor alpha  
 chain clone alpha1.4 from patient #2. It may be used, as part of  
 a method of treating rheumatoid arthritis, to raise antibodies which  
 can be administered to treat the arthritis. This therapeutic  
 approach to treatment of rheumatoid arthritis involves deletion of  
 only those T cells involved in the autoimmune response. Since these  
 comprise only a small portion of the total T cell repertoire,  
 CC eliminating these T cells should not result in significant  
 CC generalised immunosuppression. It may also be used in immunisation  
 CC to prevent the occurrence of rheumatoid arthritis.  
 SQ Sequence 108 AA;

Query Match 66.2%; Score 43; DB 7; Length 108;  
 Best Local Similarity 75.0%; Pred. No. 2.78e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 93 pdpavyql 100  
 |||||  
 QY 2 PDDAVYKYL 9

RESULT 9  
 ID W36112 standard; Protein: 112 AA.  
 AC W36112:  
 DT 19-MAY-1998 (first entry)  
 DE Human T-cell receptor alpha-chain constant region.  
 KW Human; T-cell receptor; alpha-chain constant region; antigen-specific;  
 KW immunosuppressant; humoral; cell mediated immune response; allergy;  
 KW hypersensitivity; autoimmune reaction; transplant rejection.  
 OS Homo sapiens.  
 PN WO9743411-A1.  
 PD 20-NOV-1997.  
 PF 09-MAY-1997; J01565.  
 PR 29-MAY-1996; JP-135572.  
 PR 10-MAY-1996; JP-116101.  
 PA (KIRI) KIRIN BEER KK.  
 PI Honma N, Mikayama T, Yuyama N;  
 DR WP1: 98-008880/01.  
 N-PSDB: V01420.  
 PT Immunosuppressant peptide containing T-cell receptor alpha-chain  
 PT sequence - are not antigen-specific and do not induce antibody  
 PT production  
 PS Example 10; Page 45-46; 63pp; Japanese.  
 CC The present sequence represents human T-cell receptor alpha-chain  
 CC constant region. The protein is an immunosuppressant which is not  
 CC antigen-specific and suppresses both humoral and cell-mediated immune  
 CC reactions. It can be used for treatment and/or prevention of delayed  
 CC hypersensitivity reactions, allergies and autoimmune reactions, and  
 CC inhibition of transplant rejection. The protein does not induce the  
 CC formation of antibodies against them to any significant extent.  
 SQ Sequence 112 AA;

Query Match 66.2%; Score 43; DB 28; Length 112;

Best Local Similarity 75.0%; Pred. No. 2.78e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 4 pdpavyql 11  
 |||||  
 QY 2 PDDAVYKYL 9

RESULT 10  
 ID R33270 standard; Protein: 112 AA.  
 AC R33270:  
 DT 16-JUL-1993 (first entry)  
 DE T cell receptor alpha chain clone alpha1.3.  
 KW Rheumatoid arthritis; synovial; therapy; therapeutic;  
 KW autoimmune response; variable region; mammal; immunisation.  
 OS Homo sapiens.  
 PN WO9304695-A.  
 PD 18-MAR-1993.  
 PF 27-AUG-1992; U07289.  
 PR 28-AUG-1991; US-750913.  
 PA (KIRI) UNIV PENNSYLVANIA.  
 PI (WIST-) WISTAR INST.  
 DR Weiner DB, Williams WV;  
 WP1: 93-100655/12.  
 N-PSDB: Q37077.  
 PT T-cell receptor based treatment of rheumatoid arthritis - comprises  
 administration of antibodies to T-cell receptor variable regions  
 PS Disclosure: Page 19; 110pp; English.  
 CC The sequence is that of human rheumatoid synovial T cell receptor alpha  
 CC chain clone alpha1.3 from patient #2. It may be used, as part of  
 CC a method of treating rheumatoid arthritis, to raise antibodies which  
 CC can be administered to treat the arthritis. This therapeutic  
 CC approach to treatment of rheumatoid arthritis involves deletion of  
 CC only those T cells involved in the autoimmune response. Since these  
 CC comprise only a small portion of the total T cell repertoire,  
 CC eliminating these T cells should not result in significant  
 CC generalised immunosuppression. It may also be used in immunisation  
 CC to prevent the occurrence of rheumatoid arthritis.  
 SQ Sequence 112 AA;

Query Match 66.2%; Score 43; DB 7; Length 112;  
 Best Local Similarity 75.0%; Pred. No. 2.78e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 97 pdpavyql 104  
 |||||  
 QY 2 PDDAVYKYL 9

RESULT 11  
 ID W36108 standard; Protein: 113 AA.  
 AC W36108:  
 DT 19-MAY-1998 (first entry)  
 DE Human T-cell receptor alpha-chain constant region.  
 KW Human; T-cell receptor; alpha-chain constant region; antigen-specific;  
 KW immunosuppressant; humoral; cell mediated immune response; allergy;  
 KW hypersensitivity; autoimmune reaction; transplant rejection.  
 OS Homo sapiens.  
 PN WO9743411-A1.  
 PD 20-NOV-1997.  
 PF 09-MAY-1997; J01565.  
 PR 29-MAY-1996; JP-135572.  
 PR 10-MAY-1996; JP-116101.  
 PA (KIRI) KIRIN BEER KK.  
 PI Honma N, Mikayama T, Yuyama N;  
 DR WP1: 98-008880/01.  
 N-PSDB: V01420.  
 PT Immunosuppressant peptide containing T-cell receptor alpha-chain  
 PT sequence - are not antigen-specific and do not induce antibody  
 PT production  
 PS Claim 3; Page 36; 63pp; Japanese.  
 CC The present sequence represents human T-cell receptor alpha-chain  
 CC constant region. The protein is an immunosuppressant which is not



CC antigen-specific and suppresses both humoral and cell-mediated immune  
 CC reactions. It can be used for treatment and/or prevention of delayed  
 CC hypersensitivity reactions, allergies and autoimmune reactions, and  
 CC inhibition of transplant rejection. The protein does not induce the  
 CC formation of antibodies against them to any significant extent.  
 SQ Sequence 113 AA;

Query Match 66.2%; Score 43; DB 7; Length 113;  
 Best Local Similarity 75.0%; Pred. No. 2.78e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 99 pdpavvq1 12  
 |||||  
 2 PDDAVYKTL 9

RESULT 12  
 ID R33269 standard; Protein; 114 AA.

AC R33269;  
 DT 16-JUL-1993 (first entry)  
 DE T cell receptor alpha chain clone alpha1.1/2.  
 KW Rheumatoid arthritis; synovial; therapy; therapeutic;  
 OS autoimmune response; variable region; mammal; immunisation.  
 PN WO9304695-A.  
 PD 18-MAR-1993.  
 PF 27-AUG-1992; U07289.  
 PR 28-AUG-1991; US-750913.  
 PR 06-JAN-1992; US-817912.  
 PA (UYPE-) UNIV PENNSYLVANIA.  
 PA (WIST-) WISTAR INST.  
 PI Weiner DB, Williams WV;  
 DR WPI: 93-100655/12.  
 DR N-PSDB: Q37076.  
 PT T-cell receptor based treatment of rheumatoid arthritis - comprises  
 PT administration of antibodies to T-cell receptor variable regions  
 PS Disclosure: Page 18; 110pp; English.  
 CC The sequence is that of human rheumatoid synovial T cell receptor alpha  
 CC chain clone alpha1.1/2 from patient #1. It may be used, as part of  
 CC a method of treating rheumatoid arthritis, to raise antibodies which  
 CC can be administered to treat the arthritis. This therapeutic  
 CC approach to treatment of rheumatoid arthritis involves deletion of  
 CC only those T cells involved in the autoimmune response. Since these  
 CC comprise only a small portion of the total T cell repertoire,  
 CC eliminating these T cells should not result in significant  
 CC generalised immunosuppression. It may also be used in immunisation  
 CC to prevent the occurrence of rheumatoid arthritis.  
 SQ Sequence 114 AA;

Query Match 66.2%; Score 43; DB 7; Length 114;  
 Best Local Similarity 75.0%; Pred. No. 2.78e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 99 pdpavvq1 106  
 |||||  
 2 PDDAVYKTL 9

RESULT 13  
 ID R33276 standard; Protein; 114 AA.

AC R33276;  
 DT 16-JUL-1993 (first entry)  
 DE T cell receptor alpha chain clone alpha1.7.5.  
 KW Rheumatoid arthritis; synovial; therapy; therapeutic;  
 OS autoimmune response; variable region; mammal; immunisation.  
 PN WO9304695-A.  
 PD 18-MAR-1993.  
 PF 27-AUG-1992; U07289.  
 PR 28-AUG-1991; US-750913.  
 PR 06-JAN-1992; US-817912.  
 PA (UYPE-) UNIV PENNSYLVANIA.  
 PA (WIST-) WISTAR INST.

PI Weiner DB, Williams WV;  
 DR WPI: 93-100655/12.  
 DR N-PSDB: Q37083.

PT T-cell receptor based treatment of rheumatoid arthritis - comprises  
 PT administration of antibodies to T-cell receptor variable regions  
 PS Disclosure: Page 25; 110pp; English.  
 CC The sequence is that of human rheumatoid synovial T cell receptor alpha  
 CC chain clone alpha1.7.5 from patient #7. It may be used, as part of  
 CC a method of treating rheumatoid arthritis, to raise antibodies which  
 CC can be administered to treat the arthritis. This therapeutic  
 CC approach to treatment of rheumatoid arthritis involves deletion of  
 CC only those T cells involved in the autoimmune response. Since these  
 CC comprise only a small portion of the total T cell repertoire,  
 CC eliminating these T cells should not result in significant  
 CC generalised immunosuppression. It may also be used in immunisation  
 CC to prevent the occurrence of rheumatoid arthritis.  
 SQ Sequence 114 AA;

Query Match 66.2%; Score 43; DB 7; Length 114;  
 Best Local Similarity 75.0%; Pred. No. 2.78e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 99 pdpavvq1 106  
 |||||  
 2 PDDAVYKTL 9

RESULT 14  
 ID R33275 standard; Protein; 117 AA.

AC R33275;  
 DT 16-JUL-1993 (first entry)  
 DE T cell receptor alpha chain clone alpha1.7.4.  
 KW Rheumatoid arthritis; synovial; therapy; therapeutic;  
 OS autoimmune response; variable region; mammal.  
 PN WO9304695-A.  
 PD 18-MAR-1993.  
 PF 27-AUG-1992; U07289.  
 PR 28-AUG-1991; US-750913.  
 PR 06-JAN-1992; US-817912.  
 PA (UYPE-) UNIV PENNSYLVANIA.  
 PA (WIST-) WISTAR INST.  
 PI Weiner DB, Williams WV;  
 DR WPI: 93-100655/12.  
 DR N-PSDB: Q37082.  
 PT T-cell receptor based treatment of rheumatoid arthritis - comprises  
 PT administration of antibodies to T-cell receptor variable regions  
 PS Disclosure: Page 24; 110pp; English.  
 CC The sequence is that of human rheumatoid synovial T cell receptor alpha  
 CC chain clone alpha1.7.4 from patient #4. It may be used, as part of  
 CC a method of treating rheumatoid arthritis, to raise antibodies which  
 CC can be administered to treat the arthritis. This therapeutic  
 CC approach to treatment of rheumatoid arthritis involves deletion of  
 CC only those T cells involved in the autoimmune response. Since these  
 CC comprise only a small portion of the total T cell repertoire,  
 CC eliminating these T cells should not result in significant  
 CC generalised immunosuppression. It may also be used in immunisation  
 CC to prevent the occurrence of rheumatoid arthritis.  
 SQ Sequence 117 AA;

Query Match 66.2%; Score 43; DB 7; Length 117;  
 Best Local Similarity 75.0%; Pred. No. 2.78e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 102 pdpavvq1 109  
 |||||  
 2 PDDAVYKTL 9

RESULT 15  
 ID R33273 standard; Protein; 117 AA.  
 AC R33273;  
 DT 16-JUL-1993 (first entry)

DE T cell receptor alpha chain clone alpha17.2.  
 KW Rheumatoid arthritis; synovial; therapy; therapeutic;  
 KM autoimmune response; variable region; mammal; immunisation.  
 OS Homo sapiens.  
 PN W09304695-A.  
 PD 18-MAR-1993.  
 PF 27-AUG-1992; U07289.  
 PR 28-AUG-1991; US-750913.  
 PR 06-JAN-1992; US-817912.  
 RA (UNPE-) UNIV PENNSYLVANIA.  
 RA (WIST-) WISTAR INST.  
 PI Weiner DB, Williams WV;  
 DR WPI: 93-100655/12.  
 DR N-PSDB: Q37080.  
 PT T-cell receptor based treatment of rheumatoid arthritis - comprises  
 administration of antihododies to T-cell receptor variable regions  
 PS Disclosure; Page 22; 110pp; English.  
 CC The sequence is that of human rheumatoid synovial T cell receptor alpha  
 CC chain clone alpha17.2 from patient #3. It may be used, as part of  
 CC a method of treating rheumatoid arthritis, to raise antibodies which  
 CC can be administered to treat the arthritis. This therapeutic  
 CC approach to treatment of rheumatoid arthritis involves deletion of  
 CC only those T cells involved in the autoimmune response. Since these  
 CC comprise only a small portion of the total T cell repertoire,  
 CC eliminating these T cells should not result in significant  
 CC generalised immunosuppression. It may also be used in immunisation  
 CC to prevent the occurrence of rheumatoid arthritis.  
 SO Sequence 117 AA;

## Query Match

66.2%; Score 43; DB 7; Length 117;

Best Local Similarity 75.0%; Pred. No. 2.78e+02;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 102 p0paryq1 109

QY 2 PDDAVYKL 9

Search completed: Fri Sep 11 13:01:59 1998  
 Job time: 17 secs.



T.: Attiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;  
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;  
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Moese,  
C.R.; Venter, J.C.  
Nature (1997) 390:364-370  
The complete genome sequence of the hyperthermophilic,  
sulfate-reducing archaeon *Archaeoglobus fulgidus*.  
#cross-references MUID:98049343  
#accession D69373  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
##molecule-type DNA  
##residues 1-347 #label KLE  
##cross-references GB:AE000782; TIGR:AF0988  
SUMMARY #length 347 #molecular-weight 38112 #checksum 5546

Query Match 78.5%; Score 51; DB 2; Length 347;  
Best Local Similarity 75.0%; Pred. No. 3.85e+00;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 273 PEDAVYNL 280  
1-1111111  
2 PDDAVYKL 9

RESULT 3  
ENTRY 158169 #type fragment  
TITLE semaphorin III - mouse (fragment)  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 28-Feb-1997

ACCESSIONS 158169  
REFERENCE 158169  
#authors Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;  
#journal 1 Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.  
#title Neuron (1995) 14:949-959  
#cross-references MUID:95267432  
#accession 158169  
#status preliminary; translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-666 #label RES  
##cross-references GB:L40484; NID:g703189; PID:g703190

GENETICS  
#gene SematIII  
SUMMARY #length 666 #checksum 9654

Query Match 76.9%; Score 50; DB 2; Length 666;  
Best Local Similarity 44.4%; Pred. No. 6.13e+00;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 43 HPEDNIFKL 51  
1-1111111  
1 OPDDAVYKL 9

RESULT 4  
ENTRY D49423 #type complete  
TITLE semaphorin III precursor - human  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 11-Apr-1997

ACCESSIONS D49423  
REFERENCE A49423  
#authors Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.  
#journal 1 Cell (1993) 75:1389-1399  
#title The Semaphorin genes encode a family of transmembrane and  
secreted growth cone guidance molecules.  
#accession D49423  
#status preliminary; nucleic acid sequence not shown  
##molecule-type mRNA  
##residues 1-771 #label KOL

##cross-references GB:L26081  
GENETICS  
#gene GDB:SEMA1  
##cross-references GDB:283448  
SUMMARY #length 771 #molecular-weight 88889 #checksum 6249

Query Match 76.9%; Score 50; DB 2; Length 771;  
Best Local Similarity 44.4%; Pred. No. 6.13e+00;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 149 HPEDNIFKL 157  
1-1111111  
1 OPDDAVYKL 9

RESULT 5  
ENTRY 148747 #type complete  
TITLE semaphorin D - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 28-Feb-1997

ACCESSIONS 148747  
REFERENCE 148747  
#authors Puschel, A.W.; Adams, R.H.; Betz, H.  
#journal 1 Neuron (1995) 14:941-948  
#title Murine semaphorin D/collapsin is a member of a diverse gene  
family and creates domains inhibitory for axonal extension.  
#cross-references MUID:95267431  
#accession 148747  
#status preliminary; translated from GB/EMBL/DBJ  
##molecule-type mRNA  
##residues 1-772 #label RES  
##cross-references EMBL:X85993; NID:9854329; PID:9854330

GENETICS  
#gene semD  
SUMMARY #length 772 #molecular-weight 88710 #checksum 1776

Query Match 76.9%; Score 50; DB 2; Length 772;  
Best Local Similarity 44.4%; Pred. No. 6.13e+00;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 149 HPEDNIFKL 157  
1-1111111  
1 OPDDAVYKL 9

RESULT 6  
ENTRY S59660 #type complete  
TITLE anaphase spindle elongation protein ASPL - yeast  
ORGANISM (Saccharomyces cerevisiae)  
#formal\_name Saccharomyces cerevisiae  
DATE 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 06-Feb-1998

ACCESSIONS S59660  
REFERENCE S59660  
#authors Pellman, D.; Fink, G.R.  
#journal 1 submitted to the EMBL Data Library, January 1995  
#description yeast microtubule-associated proteins required for anaphase  
spindle elongation.  
#accession S59660  
##molecule-type DNA  
##residues 1-885 #label PEL  
##cross-references EMBL:U20235; NID:g972941; PID:g972942

REFERENCE S66929  
#authors Bohm, C.; Bolotin-Fukuhara, M.; Dalgman-Fornier, B.; Dang,  
D.V.; Valens, M.  
#journal 1 submitted to the Protein Sequence Database, July 1996  
#accession S66941  
#molecule-type DNA  
#status preliminary; nucleic acid sequence not shown  
##cross-references EMBL:Z74966; NID:g1420196; PID:e252338; PID:g1420197;  
MIPS:XOR058C

|           |             |                    |             |              |
|-----------|-------------|--------------------|-------------|--------------|
|           | 23-Apr-1997 | #sequence_revision | 23-Apr-1997 | #text_change |
| SESSIONS  | 09-Sep-1997 |                    |             |              |
| S76815    |             |                    |             |              |
| REFERENCE | S74322      |                    |             |              |

Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;  
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;  
Muraki, A.; Nakazaki, N.; Naito, K.; Okumura, S.; Shimp,  
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;  
Yasuda, M.; Tabata, S.  
DNA Res. (1996) 3:109-136  
Sequence analysis of the genome of the unicellular  
cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence  
determination of the entire genome and assignment of  
potential protein-coding regions.  
#cross-references MUD:97061201  
#accession S76815  
#status Preliminary  
#molecule\_type DNA  
#residues 1-832 #label KAN  
#cross-references EMBL:D90916; NID:q1653715; PID:d1019460; PID:q1653816  
#note the nucleotide sequence was submitted to the EMBL Data  
Library, June 1996  
#length 832 #molecular-weight 92864 #checksum 8113  
SUMMARY  
Query Match 75.4%; Score 49; DB 2; Length 832;  
Best Local Similarity 44.4%; Pred. No. 9.71e+00;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 309 OPDQIFRL 317  
|||: |||:  
OY 1 OPDDAVYKL 9  
RESULT 10  
ENTRY S64506 #type complete  
TITLE protein kinase BUB1 (EC 2.7.1.-), checkpoint-associated -  
ALTERNATE\_NAMES yeast (Saccharomyces cerevisiae)  
ORGANISM protein G7342; protein YGR188c  
#formal\_name Saccharomyces cerevisiae  
DATE 17-May-1996 #sequence\_revision 17-May-1996 #text\_change  
06-Feb-1998  
ACCESSIONS S64506; A56354; S50224  
REFERENCE S64489  
#authors Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.;  
Sanchez-Perez, M.; Nombela, C.  
#submission submitted to the Protein Sequence Database, May 1996  
#accession S64506  
#molecule\_type DNA  
#residues 1-1021 #label ARR  
#cross-references EMBL:272973; NID:q1323333; PID:e243726; PID:q1323334;  
MIPS:YGR188c  
#experimental\_source strain S288C  
REFERENCE A56354  
#authors Roberts, B.T.; Farr, K.A.; Hoyt, M.A.  
#journal Mol. Cell. Biol. (1994) 14:8282-8291  
#title The Saccharomyces cerevisiae checkpoint gene BUB1 encodes a  
novel protein kinase.  
#accession A56354  
#status preliminary  
#molecule\_type DNA  
#residues 1-530, 'V', 532-1021 #label ROB  
#cross-references GB:IJ32027; NID:q475127; PID:q475128  
GENETICS SGD: BUB1  
#gene #cross-references SGD:S0003420; MIPS:YGR188c  
#map\_position 7R  
KEYWORDS autophosphorylation; cell division control; phosphoprotein;  
phosphotransferase; protein kinase  
SUMMARY #length 1021 #molecular-weight 117867 #checksum 642  
Query Match 75.4%; Score 49; DB 2; Length 1021;  
Best Local Similarity 55.6%; Pred. No. 9.71e+00;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Db 300 QSNPFYKL 308  
|||: |||:  
OY 1 OPDDAVYKL 9

RESULT 11  
ENTRY A54146 #type complete  
TITLE invasion-inducing protein Tiam-1 - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 02-Aug-1994 #sequence\_revision 02-Aug-1994 #text\_change  
10-Sep-1997  
ACCESSIONS A54146  
REFERENCE A54146  
#authors Habets, G.G.M.; Scholtes, E.H.M.; Zuydgeest, D.; van der  
Kammen, R.A.; Stam, J.C.; Berns, A.; Collard, J.C.  
Cell (1994) 77:537-549  
#title Identification of an invasion-inducing gene, Tiam-1, that  
encodes a protein with homology to GDP-GTP exchangers for  
Rho-like proteins.  
#accession A54146  
#status preliminary  
#molecule\_type mRNA  
#residues 1-1591 #label HAB  
#cross-references GB:U05245; NID:q497638; PID:q497639  
CLASSIFICATION #superfamily CDC24 homology  
FEATURE 1040-1234  
SUMMARY #domain CDC24 homology #label CD24  
#length 1591 #molecular-weight 177532 #checksum 3127  
Query Match 75.4%; Score 49; DB 2; Length 1591;  
Best Local Similarity 44.4%; Pred. No. 9.71e+00;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Db 823 OPEDIEYL 831  
|||: |||:  
OY 1 OPDDAVYKL 9  
RESULT 12  
ENTRY A33378 #type complete  
TITLE fasciclin III precursor - fruit fly (Drosophila melanogaster)  
ORGANISM #formal\_name Drosophila melanogaster  
DATE 21-Feb-1990 #sequence\_revision 21-Feb-1990 #text\_change  
23-Feb-1997  
ACCESSIONS A33378  
REFERENCE A33378  
#authors Snow, P.M.; Biebert, A.J.; Goodman, C.S.  
#journal Cell (1989) 59:313-323  
#title Fasciclin III: a novel homophilic adhesion molecule in  
Drosophila.  
#cross-references MUD:90030406  
#accession A33378  
#status preliminary  
#molecule\_type mRNA  
#residues 1-508 #label SNO  
#cross-references GB:M27813  
GENETICS FlyBase: Fasn3  
#gene #cross-references FlyBase:FBgn0000636  
KEYWORDS phosphoprotein; transmembrane protein  
SUMMARY #length 508 #molecular-weight 55883 #checksum 7642  
Query Match 73.8%; Score 48; DB 2; Length 508;  
Best Local Similarity 77.8%; Pred. No. 1.53e+01;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Db 240 OPDAAYGL 248  
|||: |||:  
OY 1 OPDDAVYKL 9  
RESULT 13  
ENTRY S27387 #type complete  
TITLE interferon alpha receptor type 1 - bovine  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change

ACCESSIONS 10-Sep-1997  
 S27387; S33770  
 REFERENCE  
 #authors Mouchel-Vielh, E.; Lutfalla, G.; Mogensen, K.E.; Ute, G.  
 #journal FEBS Lett. (1992) 313:255-259  
 #title Specific antiviral activities of the human alpha interferons are determined at the level of receptor (IFNAR) structure.  
 #accession S27387  
 #status preliminary; nucleic acid sequence not shown  
 #molecule\_type mRNA  
 #residues 1-960 #label MOU  
 #cross-references EMBL:X68443; NID:9431; PID:9432  
 #experimental\_source MDBK cells  
 REFERENCE  
 S33770  
 #authors Lim, J.K.; Langer, J.A.  
 #journal Biochim. Biophys. Acta (1993) 1173:314-319  
 #title Cloning and characterization of a bovine alpha interferon receptor.  
 #accession S33770  
 #status preliminary; nucleic acid sequence not shown  
 #molecule\_type mRNA  
 #residues 1-421 'V', 423-560 #label LIM  
 #cross-references EMBL:U06320; NID:9163187; PID:9163188  
 #experimental\_source lung  
 #title antiviral; cytokine receptor; transmembrane protein  
 KEYWORDS  
 FEATURE  
 1-24 #domain signal sequence #status predicted #label SIG  
 25-560 #product Interferon alpha receptor type I #status predicted #label MAT  
 SUMMARY #length 560 #molecular-weight 63818 #checksum 4991  
 Query Match 73.8% Score 48; DB 2; Length 560;  
 Best Local Similarity 62.5% Pred. No. 1.53e+01;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 185 PEDKLYKL 192  
 : : : : :  
 : : : : :  
 QY 2 PDDAVYKL 9

RESULT 14  
 ENTRY A27450 #type complete  
 TITLE Olfactory marker protein - rat  
 ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
 DATE 31-Dec-1988 #sequence\_revision 16-Feb-1996 #text\_change 20-Mar-1998  
 ACCESSIONS A27450; A55025  
 REFERENCE  
 #authors Rogers, K.E.; Dasgupta, P.; Gubler, U.; Grillo, M.; Khew-Goodall, Y.S.; Margolis, F.L.  
 #journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:1704-1708  
 #title Molecular cloning and sequencing of a cDNA for olfactory marker protein.  
 #cross-references MUID:87175546  
 #accession A27450  
 #molecule\_type mRNA  
 #residues 1-162 #label ROG  
 #cross-references GB:M15644; NID:9205849; PID:9205850  
 REFERENCE  
 A55025  
 #authors Sydor, W.; Tittelbaum, Z.; Blacher, R.; Sun, S.; Benz, W.; Margolis, F.L.  
 #journal Arch. Biochem. Biophys. (1986) 249:351-362  
 #title Amino acid sequence of a unique neuronal protein: rat Olfactory marker protein.  
 #accession A55025  
 #status preliminary  
 #molecule\_type protein  
 #residues 1-162 #label SYD  
 KEYWORDS acetylated amino end  
 FEATURE  
 1 #modified\_site acetylated amino end (Ala) #status experimental  
 SUMMARY #length 162 #molecular-weight 18721 #checksum 1471

Query Match 72.3% Score 47; DB 2; Length 162;  
 Best Local Similarity 44.4% Pred. No. 2.39e+01;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 46 RPAESVYRL 54  
 : : : : :  
 : : : : :  
 QY 1 QPDDAVYKL 9

RESULT 15  
 ENTRY B54261 #type complete  
 TITLE Olfactory marker protein - mouse  
 ORGANISM #formal\_name Mus musculus #common\_name house mouse  
 DATE 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 10-Sep-1997  
 ACCESSIONS B54261; I48878  
 REFERENCE  
 #authors Bulakova, O.I.; Krishna, N.S.R.; Getchell, T.V.; Margolis, F.L.  
 #journal Genomics (1994) 20:452-462  
 #title Human and rodent OMP genes: conservation of structural and regulatory motifs and cellular localization.  
 #accession B54261  
 #status preliminary  
 #molecule\_type DNA  
 #residues 1-163 #label BUI  
 #cross-references GB:U01213; NID:9457940; PID:9520741  
 REFERENCE  
 I48878  
 #authors Brown, K.A.; Sutcliffe, M.J.; Steele, K.; Brown, S.D.  
 #journal Mamm. Genome (1994) 5:11-14  
 #title Sequencing of the Olfactory Marker Protein Gene in Normal and Shaker-1 Mutant Mice.  
 #cross-references MUID:94154378  
 #accession I48878  
 #status preliminary; translated from GB/EMBL/DBJ  
 #molecule\_type DNA  
 #residues 1-163 #label RES  
 #cross-references EMBL:U02557; NID:9493516; PID:9493517  
 SUMMARY #length 163 #molecular-weight 18866 #checksum 5346

Query Match 72.3% Score 47; DB 2; Length 163;  
 Best Local Similarity 44.4% Pred. No. 2.39e+01;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 47 RPAESVYRL 55  
 : : : : :  
 : : : : :  
 QY 1 QPDDAVYKL 9

Search completed: Fri Sep 11 13:02:44 1998  
 Job time: 27 secs.

**THIS PAGE BLANK (USPTO)**





RA PELLMAN D., FINK G.R.,  
 RN SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 97279235.  
 RA VALINS M., BOHN C., DAIGMAN-FORNIER B., DANG V., BOLOTIN-FUKUHARA M.,  
 RL YEAST 13:379-390(1997).  
 CC -1- FUNCTION: REQUIRED FOR ANAPHASE SPINDLE ELONGATION.  
 DR EMBL: U20235; G972942;  
 DR EMBL: 274966; E252338;  
 DR EMBL: 270678; E234104;  
 DR SGB: 11000125; ASRL  
 SO SEQUENCE 885 AA; 101623 MW; FF008B69 CRC32;  
 DB 823 EPEHSYKL 831  
 QY 1 QPDDAVYKL 9  
 Query Match 76.9%; Score 50; DB 1; Length 885;  
 Best Local Similarity 44.4%; Pred. No. 1.45e+00;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

RESULT 3  
 ID RS4E CANAL STANDARD; PRT; 262 AA.  
 AC P47837;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE 40S RIBOSOMAL PROTEIN S4 (S7).  
 GN RPS7  
 OS CANDIDA ALBICANS (YEAST).  
 OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SG7.243;  
 RA DELBUECK S., SONNEBORN A., GERADS M., GRABLOWITZ A.H., ERNST J.F.,  
 RL YEAST 13:1199-1210(1997).  
 CC -1- SIMILARITY: BELONGS TO THE S4E FAMILY OF RIBOSOMAL PROTEINS.  
 DR EMBL: U37009; G1051260;  
 DR PROSITE: PS00528; RIBOSOMAL\_S4E; 1.  
 KW RIBOSOMAL PROTEIN.  
 SO SEQUENCE 262 AA; 29204 MW; CE29056D CRC32;  
 DB 116 AEEAVYKL 123  
 QY 2 PDDAVYKL 9  
 Query Match 75.4%; Score 49; DB 1; Length 262;  
 Best Local Similarity 62.5%; Pred. No. 2.46e+00;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

RESULT 4  
 ID PIGR RABIT STANDARD; PRT; 773 AA.  
 AC P01832;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (PIGR) (CONTAINS:  
 DE SECRETORY COMPONENT).  
 GN PIGR  
 OS ORCTOLAGUS CUNICULUS (RABBIT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; LAGOMORPHA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84142246.  
 RA MOSTOV K.E., FRIEDLANDER M., BLOBEL G.,  
 RL NATURE 308:37-43(1984).  
 [2]  
 RN SEQUENCE OF 87-114 AND 410-428.  
 RP STRAIN-S288C;  
 RX MEDLINE: 88228032.

RA FRUTIGER S., HUGHES G.J., HANLY W.C., JATON J.-C.;  
 RL J. BIOL. CHEM. 263:8120-8125(1988).  
 CC -1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE  
 CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN  
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.  
 CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE  
 CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE  
 CC TRANSMEMBRANE SEGMENT.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.  
 CC -1- THE SEQUENCE SHOWN IS THAT OF ALLOTYP E T62.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
 CC FIVE V-LIKE DOMAINS.  
 DR EMBL: X00412; G1596;  
 DR EMBL: K01291; G155106;  
 DR PIR: A02111; QRRSG.  
 DR PIR: A28077; A28077.  
 DR HSSE: P07006; IJBL.  
 KW IMMUNOGLOBULIN FOLD; REPEAT; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;  
 KM POLYMORPHISM.  
 FT SIGNAL 1 18  
 FT CHAIN 19 773  
 FT CHAIN 19 615  
 FT DOMAIN 19 647  
 FT TRANSMEM 648 670  
 FT DOMAIN 671 773  
 FT DOMAIN 30 136  
 FT DOMAIN 137 243  
 FT DOMAIN 244 350  
 FT DOMAIN 351 456  
 FT DOMAIN 457 558  
 FT DISULFID 46 115  
 FT DISULFID 155 225  
 FT DISULFID 260 324  
 FT DISULFID 369 438  
 FT DISULFID 478 538  
 FT CARBOHYD 88 88  
 FT CARBOHYD 108 108  
 FT CARBOHYD 418 418  
 FT VARIANT 88 88  
 FT VARIANT 94 94  
 FT VARIANT 101 108  
 FT VARIANT 110 110  
 SO SEQUENCE 773 AA; 83886 MW; 79840D1F CRC32;  
 DB 137 EPDDVYKL 144  
 QY 1 QPDDAVYKL 8  
 Query Match 75.4%; Score 49; DB 1; Length 773;  
 Best Local Similarity 75.0%; Pred. No. 2.46e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 5  
 ID BUB1 YEAST STANDARD; PRT; 1021 AA.  
 AC P41695;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CHECKPOINT SERINE/THREONINE-PROTEIN KINASE BUB1 (EC 2.7.1.-).  
 GN BUB1 OR YGR188C OR G7542.  
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE: 95059057.  
 RA ROBERTS B.T., FARR K.A., HOYT M.A.,  
 RL MOL. CELL. BIOL. 14:8282-8291(1994).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-S288C;  
 RX MEDLINE: 97279231.

RA ARROYO J., GARCIA-GONZALEZ M., GARCIA-SAEZ M.I., SANCHEZ-PEREZ M.,  
RA NOMBELA C.;  
RL YEAST 13:357-363(1997).  
CC -1- FUNCTION: INVOLVED IN CELL CYCLE CHECKPOINT ENFORCEMENT. CATALYZES  
CC THE PHOSPHORYLATION OF HUB3 AND ITS AUTOPHOSPHORYLATION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- PTM: AUTOPHOSPHORYLATED.  
CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-  
CC PROTEIN KINASES.  
CC -1- SIMILARITY: SOME, IN THE N-TERMINUS WITH THE N-TERMINUS OF MAD3.  
CC -1- SIMILARITY: IN THE N-TERMINUS, WITH YEAST YJL013C.  
DR EMBL: L32027; G475128; -  
DR EMBL: Z72973; E243726; -  
DR EMBL: X99074; E252632; -  
DR SGD: L0000196; BUB1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
KM TRANSFERASE: SERINE/THREONINE-PROTEIN KINASE: ATP-BINDING;  
KM CELL CYCLE: PHOSPHORYLATION: NUCLEAR PROTEIN.  
FT DOMAIN 705 1021  
FT NP\_BIND 711 719  
FT BINDING 733 733  
FT ACT\_SITE 833 833  
FT MUTAGEN 733 733  
FT CONFLICT 531 531  
SQ SEQUENCE 1021 AA; 117868 MW; C9532F44 CRC32;  
D -> V (IN REF. 1).  
Query Match  
Best Local Similarity 75.4%; Score 49; DB 1; Length 1021;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Db 300 QSNPPYKYL 308  
QY 1 QPDDAVYKYL 9  
RESULT 6  
ID TIAM\_HUMAN STANDARD; PRT; 1591 AA.  
AC Q13009;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE T-LYMPHOMA INVASION AND METASTASIS INDUCING PROTEIN 1 (TIAM1 PROTEIN).  
GN TIAM1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX HABETS G.G., DER KAMMEN R.A., STAM J.C., MICHELS F., COLLARD J.G.;  
RX HABETS G.G., DER KAMMEN R.A., STAM J.C., MICHELS F., COLLARD J.G.;  
RL ONCOGENE 10:1371-1376(1995).  
RN [2]  
RP INTERACTIONS WITH RAC.  
RC TISSUE-BRAIN;  
RX MICHELS F., HABETS G.G., STAM J.C., DER KAMMEN R.A., COLLARD J.G.;  
RL NATURE 375:338-340(1995).  
RN [3]  
RP MAPPING.  
RX MEDLINE: 95254877.  
RA HABETS G.G., DER KAMMEN R.A., JENKINS N.A., GILBERT D.J.,  
RA COPELAND N.G., HAGEMEIJER A., COLLARD J.G.;  
RL CYTOGENET. CELL GENER. 70:48-51(1995).  
CC -1- FUNCTION: MAY FUNCTION IN CELLULAR SIGNALING PROCESSES PRESUMABLY  
CC BY ACTIVATION OF A RHO-LIKE GTPASE THAT REGULATES THE CYTOSKELETAL  
CC ORGANIZATION.  
CC -1- TISSUE SPECIFICITY: FOUND IN VIRTUALLY ALL ANALYSED TUMOR CELL  
CC LINES INCLUDING B- AND T- LYMPHOMAS, NEUROBLASTOMAS, MELANOMAS AND  
CC CARCINOMAS.  
CC -1- SIMILARITY: TO OTHER GUANINE-NUCLEOTIDE RELEASING FACTORS OF THE

CC CDC24 FAMILY.  
CC -1- SIMILARITY: CONTAINS A PH DOMAIN.  
DR EMBL: U16286; G897557; -  
DR PROSITE: PS00741; GDS\_CDC24; 1.  
DR PROSITE: PS50003; PH\_DOMAIN; 1.  
KM GUANINE-NUCLEOTIDE RELEASING FACTOR.  
FT DOMAIN ? ?  
FT DOMAIN 595 598  
FT DOMAIN 1445 1449  
SQ SEQUENCE 1591 AA; 177637 MW; 758BCB0E CRC32;  
Query Match  
Best Local Similarity 75.4%; Score 49; DB 1; Length 1591;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Db 823 QPEDDIYEL 831  
QY 1 QPDDAVYKYL 9  
RESULT 7  
ID TIAM\_MOUSE STANDARD; PRT; 1591 AA.  
AC Q60610;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE T-LYMPHOMA INVASION AND METASTASIS INDUCING PROTEIN 1 (TIAM1 PROTEIN).  
GN TIAM1 OR TIAM-1.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C; TISSUE-BRAIN;  
RX MEDLINE: 94243921.  
RA HABETS G.G.M., SCHOLTES E.H.M., ZUYDEBEST D., DER KAMMEN R.A.,  
RA STAM J.C., COLLARD J.G.;  
RL CELL 77:537-549(1994).  
CC -1- FUNCTION: MAY FUNCTION IN CELLULAR SIGNALING PROCESSES PRESUMABLY  
CC BY ACTIVATION OF A RHO-LIKE GTPASE THAT REGULATES THE CYTOSKELETAL  
CC ORGANIZATION.  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN AND TESTIS AND AT  
CC LOW OR MODERATE LEVELS IN ALMOST ALL OTHER NORMAL TISSUES. FOUND  
CC IN VIRTUALLY ALL ANALYSED TUMOR CELL LINES INCLUDING B- AND T-  
CC LYMPHOMAS, NEUROBLASTOMAS, MELANOMAS AND CARCINOMAS.  
CC -1- SIMILARITY: TO OTHER GUANINE-NUCLEOTIDE RELEASING FACTORS OF THE  
CC CDC24 FAMILY.  
DR EMBL: U05245; G497639; -  
DR MGD: MGI:103306; TIAM1.  
DR PROSITE: PS00741; GDS\_CDC24; 1.  
DR PROSITE: PS50003; PH\_DOMAIN; 1.  
KM GUANINE-NUCLEOTIDE RELEASING FACTOR.  
FT DOMAIN ? ?  
FT DOMAIN 595 598  
FT DOMAIN 1445 1449  
SQ SEQUENCE 1591 AA; 177532 MW; 0220ECCC CRC32;  
Query Match  
Best Local Similarity 75.4%; Score 49; DB 1; Length 1591;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Db 823 QPEDDIYEL 831  
QY 1 QPDDAVYKYL 9  
RESULT 8  
ID Y4VH\_RHISN STANDARD; PRT; 218 AA.  
AC Q53216;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

| FT                    | MOD_HYD   | 300       | 300          | POTENTIAL.                            |
|-----------------------|---|-----------|--------------|---------------------------------------|
| FT                    | CARBOHYD  | 300       | 300          | POTENTIAL.                            |
| FT                    | MOD_RES   | 382       | 382          | PHOSPHORYLATION (POTENTIAL).          |
| FT                    | MOD_RES   | 459       | 459          | PHOSPHORYLATION (POTENTIAL).          |
| SO                    | SEQUENCE  | 508 AA;   | 55883 MM;    | C9417AFB CRC32;                       |
| Query Match           |   |           |              |                                       |
| Best Local Similarity |   | 73.8%;    | Score 48;    | DB 1; Length 508;                     |
| Matches               |   | 7;        | Conservative | 0; Mismatches 2; Indels 0; Gaps 0;    |
| Db                    | 240 QPDAVYGL 248  |           |              |                                       |
| Oy                    | 1 QPDAVYKL 9  |           |              |                                       |
| RESULT 10             |   |           |              |                                       |
| ID                    | INRL_SHEEP  | STANDARD; | PRT;         | 560 AA.                               |
| AC                    | Q28589; Q95206;   |           |              |                                       |
| DT                    | 01-NOV-1997 (REL. 35, CREATED)  |           |              |                                       |
| DT                    | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)                         |           |              |                                       |
| DT                    | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)                         |           |              |                                       |
| DE                    | INTERFERON-ALPHA/BETA RECEPTOR ALPHA PRECURSOR (IFN-ALPHA-REC)        |           |              |                                       |
| DE                    | (INTERFERON ALPHA/BETA RECEPTOR-1).                                   |           |              |                                       |
| GN                    | IFNAR1 OR IFNAR.  |           |              |                                       |
| OS                    | OVIS ARIES (SHEEP).   |           |              |                                       |
| OC                    | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;        |           |              |                                       |
| CC                    | EUTHERIA; ARTIODACTYLA.   |           |              |                                       |
| RN                    | [1]   |           |              |                                       |
| RP                    | SEQUENCE FROM N.A.  |           |              |                                       |
| RA                    | KALUZOVA S., FISCHER P.A., KALUZOVA M., SHELDRIK E.L., FLINT A.P.F.;  |           |              |                                       |
| RL                    | SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.                  |           |              |                                       |
| RP                    | (2)   |           |              |                                       |
| RP                    | SEQUENCE FROM N.A.  |           |              |                                       |
| RC                    | TISSUE-ENDOMETRIUM;   |           |              |                                       |
| RA                    | HAN C.S., MATHIALAGAN N., KLEMAN S.W., ROBERTS R.M.;                  |           |              |                                       |
| RL                    | SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.                  |           |              |                                       |
| CC                    | -1- FUNCTION: RECEPTOR FOR INTERFERONS ALPHA AND BETA. BINDING TO TO  |           |              |                                       |
| CC                    | TYPE I IFNS TRIGGERS TYROSINE PHOSPHORYLATION OF A NUMBER OF          |           |              |                                       |
| CC                    | PROTEINS INCLUDING JAKS, TYR2, STAT PROTEINS AND IFN-R ALPHA- AND     |           |              |                                       |
| CC                    | BETA-SUBUNITs THEMSELVES.   |           |              |                                       |
| CC                    | -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.                    |           |              |                                       |
| CC                    | -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.         |           |              |                                       |
| CC                    | -1- SIMILARITY: BELONGS TO THE CLASS II CYTOKINE FAMILY OF RECEPTORS. |           |              |                                       |
| DR                    | EMBL; X95939; E224016; -  |           |              |                                       |
| DR                    | EMBL; U65978; G153002; -  |           |              |                                       |
| KW                    | RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.                        |           |              |                                       |
| FT                    | SIGNAL  | 1         | 24           | BY SIMILARITY.                        |
| FT                    | CHAIN   | 25        | 560          | INTERFERON-ALPHA/BETA RECEPTOR ALPHA. |
| FT                    | CHAIN   |           |              |                                       |
| FT                    | DOMAIN  | 25        | 437          | EXTRACELLULAR (POTENTIAL).            |
| FT                    | TRANSMEM  | 438       | 458          | POTENTIAL.                            |
| FT                    | DOMAIN  | 459       | 560          | CYTOPLASMIC (POTENTIAL).              |
| FT                    | DISULEID  | 76        | 84           | BY SIMILARITY.                        |
| FT                    | DISULEID  | 199       | 220          | BY SIMILARITY.                        |
| FT                    | CARBOHYD  | 47        | 47           | POTENTIAL.                            |
| FT                    | CARBOHYD  | 35        | 55           | POTENTIAL.                            |
| FT                    | CARBOHYD  | 85        | 85           | POTENTIAL.                            |
| FT                    | CARBOHYD  | 108       | 108          | POTENTIAL.                            |
| FT                    | CARBOHYD  | 109       | 109          | POTENTIAL.                            |
| FT                    | CARBOHYD  | 172       | 172          | POTENTIAL.                            |
| FT                    | CARBOHYD  | 222       | 222          | POTENTIAL.                            |
| FT                    | CARBOHYD  | 285       | 285          | POTENTIAL.                            |
| FT                    | CARBOHYD  | 313       | 313          | POTENTIAL.                            |
| FT                    | CARBOHYD  | 359       | 359          | POTENTIAL.                            |
| FT                    | CARBOHYD  | 377       | 377          | POTENTIAL.                            |
| FT                    | CARBOHYD  | 434       | 434          | POTENTIAL.                            |
| FT                    | CONFLICT  | 9         | 10           | TL -> LT (IN REF. 2).                 |
| FT                    | CONFLICT  | 322       | 323          | KE -> RK (IN REF. 2).                 |
| FT                    | CONFLICT  | 415       | 416          | RW -> WR (IN REF. 2).                 |
| FT                    | CONFLICT  | 522       | 522          | A -> D (IN REF. 2).                   |
| SO                    | SEQUENCE  | 560 AA;   | 63918 MM;    | 33418DCG CRC32;                       |
| Query Match           |   |           |              |                                       |
| Best Local Similarity |   | 73.8%;    | Score 48;    | DB 1; Length 560;                     |
| Matches               |   | 73.8%;    | Score 48;    | DB 1; Length 560;                     |
| Best Local Similarity |   | 62.5%;    | Score 48;    | DB 1; Length 560;                     |
| Best Local Similarity |   | 62.5%;    | Score 48;    | DB 1; Length 560;                     |

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 185 PEDIKYKL 192  
|:|:|:|

OY 2 PDDAVYKL 9

RESULT 11  
ID INRL BOVIN STANDARD; PRT; 560 AA.  
AC 004780;  
DT 01-OCT-1993 (REL. 27, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE INTERFERON-ALPHA/BETA RECEPTOR ALPHA CHAIN PRECURSOR (IFN-ALPHA-REC).  
GN IFNARI OR IFNAR.  
OS BOS TAURUS (BOVINE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; ARTIODACTYLA.  
[1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG;  
RX MEDLINE: 93076908.  
RA MOUCHEL-VIELEH E., MOGENSEN K.E., UZE G.;  
RL FEBS LETT. 313:255-259(1992).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 93305725.  
RA LIM J.-K., LANGER J.A.;  
RL BIOCHIM. BIOPHYS. ACTA 1173:314-319(1993).  
CC -1- FUNCTION: RECEPTOR FOR INTERFERON ALPHA AND BETA. BINDING TO TO  
TYPE I IFNS TRIGGERS TYROSINE PHOSPHORYLATION OF A NUMBER OF  
PROTEINS INCLUDING JAKS, TYK2, STAT PROTEINS AND IFN-R ALPHA- AND  
BETA-SUBUNITS THEMSELVES.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
CC -1- SIMILARITY: BELONGS TO THE CLASS II CYTOKINE FAMILY OF RECEPTORS.  
DR EMBL: X68443; G433; -;  
DR EMBL: L06320; G163188; -;  
DR PIR: S33770; S33770.  
DR PIR: S27387; S27387.  
KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.  
FT SIGNAL 1 24  
FT CHAIN 25 560  
FT DOMAIN 25 437  
FT TRANSMEM 438 437  
FT DOMAIN 459 560  
FT DISULFID 76 84  
FT DISULFID 199 220  
FT CARBOHYD 47 47  
FT CARBOHYD 55 55  
FT CARBOHYD 85 85  
FT CARBOHYD 109 109  
FT CARBOHYD 172 172  
FT CARBOHYD 254 254  
FT CARBOHYD 313 313  
FT CARBOHYD 377 377  
FT CARBOHYD 434 434  
FT CONFLICT 422 422  
SQ SEQUENCE 560 AA; 63818 MW; 44D98FDF CRC32;

Query Match 73.8%; Score 48; DB 1; Length 560;  
Best Local Similarity 62.5%; Pred. No. 4.14e+00;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 185 PEDIKYKL 192  
|:|:|:|

OY 2 PDDAVYKL 9

RESULT 12  
ID OMP\_RAT STANDARD; PRT; 162 AA.  
AC P08523;

DT 01-AUG-1988 (REL. 08, CREATED)  
DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE OLFACTORY MARKER PROTEIN (OLFACTORY NEURONAL SPECIFIC PROTEIN).  
GN OMP.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87175546.  
RA ROGERS K.E., DASGUPTA P., GUBLER U., GRILLO M., KHEW-GOODALL Y.S.,  
RA MAROLIS F.L.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 84:1704-1708(1987).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90046838.  
RA DANCIGER E., METTLING C., VIDAL M., MORRIS R., MARGOLIS F.L.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 86:8565-8569(1989).  
[3]  
RP SEQUENCE.  
RX MEDLINE: 86321994.  
RA SYDOR W., TEITELBAUM Z., BLACHER R., SUN S., BENZ W., MARGOLIS F.L.;  
RL ARCH. BIOCHEM. BIOPHYS. 249:351-362(1986).  
CC -1- TISSUE SPECIFICITY: UNIQUELY ASSOCIATED WITH MATURE OLFACTORY  
RECEPTOR NEURONS.  
CC EMBL: M15644; G205850; -;  
DR EMBL: M26926; G205852; -;  
DR PIR: A27450; A27450.  
DR PIR: A55025; A55025.  
KW NEURONE; OLFACTION; ACETYLATION.  
FT INIT\_MET 0 0  
FT MOD\_RES 1 1  
SQ SEQUENCE 162 AA; 18721 MW; 11FD9FB8 CRC32;

Query Match 72.3%; Score 47; DB 1; Length 162;  
Best Local Similarity 44.4%; Pred. No. 6.91e+00;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 46 RPAESYRL 54  
|:|:|:|

OY 1 QPDDAVYKL 9

RESULT 13  
ID G6PD\_SCHPO STANDARD; PRT; 447 AA.  
AC 000091;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD) (FRAGMENT).  
GN ZWFI OR SPAC312.18.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMICETES.  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA BADCOCK K., CHURCHER C.M., WOOD V., BARRELL B.G., RAJANDREAM M.A.;  
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) -  
D-GLUCONO-DELTA-LACTONE 6-PHOSPHATE + NADPH.  
CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.  
DR PROSITE: PS00069; G6P DEHYDROGENASE; 1.  
KW OXIDOREDUCTASE; NADP: GLUCOSE METABOLISM.  
FT ACT\_SITE 189 189  
FT NON\_TER 447 447  
SQ SEQUENCE 447 AA; 50926 MW; 59CBE54F CRC32;

Query Match 72.3%; Score 47; DB 1; Length 447;  
Best Local Similarity 71.4%; Pred. No. 6.91e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 375 OPDEALY 381  
 1111111  
 OY 1 OPDAVY 7

RESULT 14  
 ID YVHF\_BACSU STANDARD: PRT: 513 AA.  
 AC P71067.  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE PUTATIVE L-LACTATE PERMEASE YVHF (FRAGMENT).  
 GN YVHF.  
 OS BACILLUS SUBTILIS.  
 OC PROAROT; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168.  
 RA FABRET C., QUENTIN Y., CHAPAL N., GUISEPI A., HAIECH J., DENIZOT F.;  
 RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: MAY PLAY A ROLE IN L-LACTATE TRANSPORT.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE LIDP FAMILY OF TRANSPORTERS.  
 DR EMBL: 271928: E238672: -.  
 DR SUBTILIST; BG11875; YVHF.  
 KM HYPOTHETICAL PROTEIN; TRANSPORT; TRANSMEMBRANE.  
 FT NON\_TER 1  
 FT TRANSMEM 23 43 POTENTIAL.  
 FT TRANSMEM 81 101 POTENTIAL.  
 FT TRANSMEM 107 127 POTENTIAL.  
 FT TRANSMEM 144 164 POTENTIAL.  
 FT TRANSMEM 170 190 POTENTIAL.  
 FT TRANSMEM 199 219 POTENTIAL.  
 FT TRANSMEM 254 274 POTENTIAL.  
 FT TRANSMEM 331 351 POTENTIAL.  
 FT TRANSMEM 369 389 POTENTIAL.  
 FT TRANSMEM 398 418 POTENTIAL.  
 FT TRANSMEM 456 476 POTENTIAL.  
 FT TRANSMEM 492 512 POTENTIAL.  
 SQ SEQUENCE 513 AA; 54395 MW; AAB47105 CRC32;

Query Match 72.3%; Score 47; DB 1; Length 513;  
 Best Local Similarity 75.0%; Pred. No. 6.91e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 321 PIEAVYKL 328  
 1111111  
 OY 2 PDDAVYKL 9

RESULT 15  
 ID FAS\_BOVIN STANDARD: PRT: 2211 AA.  
 AC Q28107; Q28108;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE COAGULATION FACTOR V PRECURSOR (ACTIVATED PROTEIN C COFACTOR).  
 GN F5.  
 OS BOS TAURUS (BOVINE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; ARTIODACTYLIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 92147638.  
 RA GUINTO E.R., ESMON C.T., MANN K.G., MACGILLIVRAY R.T.;  
 RE J. BIOL. CHEM. 267:2971-2978(1992).  
 CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES  
 WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.  
 CC -1- SUBUNIT: FACTOR VA IS COMPOSED OF AN HEAVY CHAIN AND OF A LIGHT  
 CHAIN NONCOVALENTLY BOUND. THE INTERACTION BETWEEN THE TWO CHAINS  
 IS CALCIUM-DEPENDENT.

CC -1- DOMAIN: DOMAIN B CONTAINS 29.5 X 9 AA TANDEM REPEATS, AND 2 X 17  
 CC AA REPEATS.  
 CC -1- PTM: THROMBIN ACTIVATES FACTOR V PROTEOLITICALLY TO THE ACTIVE  
 CC COFACTOR, FACTOR V(A) (FORMATION OF A HEAVY CHAIN AT THE N-  
 CC TERMINUS AND A LIGHT CHAIN AT THE C-TERMINUS).  
 CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF  
 CC 2 PLASTOCYANIN-LIKE REPEATS.  
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.  
 CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.  
 DR EMBL: M81440; G163038; -.  
 DR EMBL: M81441; G163040; -.  
 DR PROSITE: PS00079; MULTICOOPER\_OXIDASE1. 2.  
 DR PROSITE: PS01285; FAS5C\_1. 2.  
 DR PROSITE: PS01286; FAS5C\_2. 2.  
 DR BLOOD COAGULATION; PLASMA; GLYCOPROTEIN; CALCIUM; SIGNAL; ZMOGEN;  
 KW REPEAT.  
 FT SIGNAL 1 28  
 FT CHAIN 29 2211  
 FT CHAIN 29 741  
 FT PEPTIDE 742 1564

CHAIN 1565 2211  
 DOMAIN 30 327  
 DOMAIN 30 193  
 DOMAIN 203 327  
 DOMAIN 348 686  
 DOMAIN 348 525  
 DOMAIN 535 686  
 DOMAIN 696 1564  
 SIMILAR 899 915  
 DOMAIN 1124 1151  
 REPEAT 1124 1137  
 REPEAT 1138 1151  
 DOMAIN 1188 1453

REPEAT 1188 1196  
 REPEAT 1197 1205  
 REPEAT 1206 1214  
 REPEAT 1215 1223  
 REPEAT 1223 1232  
 REPEAT 1233 1241  
 REPEAT 1242 1250  
 REPEAT 1251 1259  
 REPEAT 1260 1268  
 REPEAT 1269 1277  
 REPEAT 1278 1286  
 REPEAT 1287 1295  
 REPEAT 1296 1304  
 REPEAT 1305 1313  
 REPEAT 1314 1322  
 REPEAT 1323 1331  
 REPEAT 1332 1340  
 REPEAT 1341 1349  
 REPEAT 1349 1358  
 REPEAT 1359 1367  
 REPEAT 1368 1376  
 REPEAT 1377 1385  
 REPEAT 1386 1394  
 REPEAT 1395 1403  
 REPEAT 1404 1412  
 REPEAT 1413 1421  
 REPEAT 1422 1430  
 REPEAT 1431 1439  
 REPEAT 1439 1447  
 REPEAT 1444 1453  
 REPEAT 1453 1461  
 DOMAIN 1569 1890  
 DOMAIN 1569 1738  
 DOMAIN 1748 1890  
 DOMAIN 1894 2048  
 DOMAIN 2053 2208  
 SITE 741 742  
 SITE 1034 1035  
 SITE 1564 1565

POTENTIAL.  
 COAGULATION FACTOR V.  
 HEAVY CHAIN (BY SIMILARITY).  
 ACTIVATION PEPTIDE (CONNECTING REGION)  
 (BY SIMILARITY).  
 LIGHT CHAIN (BY SIMILARITY).  
 F5/8 TYPE A 1.  
 PLASTOCYANIN-LIKE 1.  
 PLASTOCYANIN-LIKE 2.  
 F5/8 TYPE A 2.  
 PLASTOCYANIN-LIKE 3.  
 PLASTOCYANIN-LIKE 4.  
 B.  
 TO 17 AA REPEATS IN HUMAN FAS.  
 2 X 14 AA TANDEM REPEATS.  
 1.  
 2.  
 30 X 9 AA TANDEM REPEATS OF [AS]-L-S-P-  
 D-[LP]-[GS]-O-[TE] (APPROXIMATE).  
 1.  
 2.  
 3.  
 4.  
 5.  
 6.  
 7.  
 8.  
 9.  
 10.  
 11.  
 12.  
 13.  
 14.  
 15.  
 16.  
 17.  
 18.  
 19.  
 20.  
 21.  
 22.  
 23.  
 24.  
 25.  
 26.  
 27.  
 28.  
 29 (PARTIAL).  
 30.  
 F5/8 TYPE A 3.  
 PLASTOCYANIN-LIKE 5.  
 PLASTOCYANIN-LIKE 6.  
 F5/8 TYPE C 1.  
 F5/8 TYPE C 2.  
 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).  
 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).  
 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).

|    |          |      |            |                             |
|----|----------|------|------------|-----------------------------|
| FT | DISULFID | 167  | 193        | PROBABLE.                   |
| FT | DISULFID | 499  | 525        | PROBABLE.                   |
| FT | DISULFID | 1712 | 1738       | PROBABLE.                   |
| FT | DISULFID | 1894 | 2048       | BY SIMILARITY.              |
| FT | DISULFID | 2053 | 2208       | BY SIMILARITY.              |
| FT | CARBOHYD | 225  | 225        | POTENTIAL.                  |
| FT | CARBOHYD | 239  | 239        | POTENTIAL.                  |
| FT | CARBOHYD | 297  | 297        | POTENTIAL.                  |
| FT | CARBOHYD | 382  | 382        | POTENTIAL.                  |
| FT | CARBOHYD | 460  | 460        | POTENTIAL.                  |
| FT | CARBOHYD | 553  | 553        | POTENTIAL.                  |
| FT | CARBOHYD | 587  | 587        | POTENTIAL.                  |
| FT | CARBOHYD | 745  | 745        | POTENTIAL.                  |
| FT | CARBOHYD | 756  | 756        | POTENTIAL.                  |
| FT | CARBOHYD | 774  | 774        | POTENTIAL.                  |
| FT | CARBOHYD | 780  | 780        | POTENTIAL.                  |
| FT | CARBOHYD | 902  | 902        | POTENTIAL.                  |
| FT | CARBOHYD | 952  | 952        | POTENTIAL.                  |
| FT | CARBOHYD | 964  | 964        | POTENTIAL.                  |
| FT | CARBOHYD | 1044 | 1044       | POTENTIAL.                  |
| FT | CARBOHYD | 1053 | 1053       | POTENTIAL.                  |
| FT | CARBOHYD | 1062 | 1062       | POTENTIAL.                  |
| FT | CARBOHYD | 1071 | 1071       | POTENTIAL.                  |
| FT | CARBOHYD | 1078 | 1078       | POTENTIAL.                  |
| FT | CARBOHYD | 1094 | 1094       | POTENTIAL.                  |
| FT | CARBOHYD | 1451 | 1451       | POTENTIAL.                  |
| FT | CARBOHYD | 1490 | 1490       | POTENTIAL.                  |
| FT | CARBOHYD | 1550 | 1550       | POTENTIAL.                  |
| FT | CARBOHYD | 1690 | 1690       | POTENTIAL.                  |
| FT | CARBOHYD | 1839 | 1839       | POTENTIAL.                  |
| FT | CARBOHYD | 1997 | 1997       | POTENTIAL.                  |
| FT | CARBOHYD | 2196 | 2196       | POTENTIAL.                  |
| FT | VARIANT  | 587  | 592        | NETLPA -> T (IN VARIANT 2). |
| SO | SEQUENCE | 2211 | AA: 248981 | MW: 9B017C5C CRC32;         |

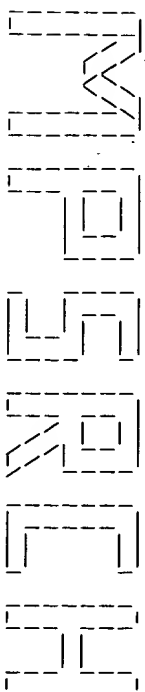
Query Match 72.3%; Score 47; DB 1; Length 2211;  
Best Local Similarity 71.4%; Pred. No. 6,91e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 1597 PEDTVYK 1603  
OY 2 PDDAVYK 8

Search completed: Fri Sep 11 13:03:07 1998  
Job time : 6 secs.

**THIS PAGE BLANK (USPTO)**





(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*  
MPSrch\_PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Sep 11 13:03:25 1998; Maspar time 4.09 Seconds  
Tabular output not generated. 92.584 Million cell updates/sec

Title: >US-08-452-843-7  
Description: (1-9) from US08452843.pep  
Perfect score: 65  
Sequence: 1 QPDDAVYKL 9

Scoring table:  
PAM 150  
Gap 15

Searched: 140555 segs, 42109429 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

sptrembl6  
1:sp:func1 2:sp:human 3:sp:invertebrate 4:sp:mammal  
5:sp:mhc 6:sp:organelle 7:sp:phage 8:sp:plant  
9:sp:bacteria 10:sp:rodent 11:sp:virus 12:sp:vertebrate  
13:sp:unclassified

Statistics: Mean 22.797; Variance 25.825; scale 0.883

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description            | Pred. No. |
|------------|-------|-------------|--------|-------|------------------------|-----------|
| 1          | 51    | 78.5        | 347    | 9     | IMMUNOGENIC PROTEIN (B | 1.58e+00  |
| 2          | 51    | 78.5        | 437    | 2     | HYPOTHEICAL 48.5 KD P  | 1.58e+00  |
| 3          | 50    | 76.9        | 666    | 10    | SEMAPHORIN D (SEMAPHOR | 2.66e+00  |
| 4          | 50    | 76.9        | 771    | 2     | SEMAPHORIN-III.        | 2.66e+00  |
| 5          | 50    | 76.9        | 772    | 10    | COLLAPIN-1.            | 2.66e+00  |
| 6          | 50    | 76.9        | 772    | 10    | SEMAPHORIN III/COLLAPS | 2.66e+00  |
| 7          | 50    | 76.9        | 772    | 10    | SEMAPHORIN D.          | 2.66e+00  |
| 8          | 50    | 76.9        | 772    | 10    | SEMAPHORIN D.          | 2.66e+00  |
| 9          | 49    | 75.4        | 832    | 9     | MANNOSE-6-PHOSPHATE A  | 2.66e+00  |
| 10         | 48    | 73.8        | 1307   | 3     | HYPOTHEICAL 92.9 KD P  | 4.43e+00  |
| 11         | 47    | 72.3        | 1307   | 3     | T22C1.10.              | 7.32e+00  |
| 12         | 47    | 72.3        | 163    | 10    | OLFACTORY MARKER PROTE | 1.20e+01  |
| 13         | 47    | 72.3        | 317    | 8     | GLUCOSE-6-PHOSPHATE 1- | 1.20e+01  |
| 14         | 47    | 72.3        | 465    | 8     | GLUCOSE-6-PHOSPHATE 1- | 1.20e+01  |
| 15         | 47    | 72.3        | 492    | 8     | GLUCOSE-6-PHOSPHATE 1- | 1.20e+01  |
| 16         | 47    | 72.3        | 522    | 3     | GLUCOSE-6-PHOSPHATE 1- | 1.20e+01  |
| 17         | 47    | 72.3        | 574    | 8     | GLUCOSE-6-PHOSPHATE 1- | 1.20e+01  |
| 18         | 47    | 72.3        | 577    | 8     | GLUCOSE-6-PHOSPHATE 1- | 1.20e+01  |
| 19         | 47    | 72.3        | 604    | 8     | GLUCOSE-6-PHOSPHATE 1- | 1.20e+01  |
| 20         | 46    | 70.8        | 1215   | 9     | DNA FOR SEROTYPE B CAP | 1.20e+01  |
|            |       |             | 387    | 9     | CARBOXYNOSPHEMIDINE D  | 1.96e+01  |
|            |       |             |        |       | P73562                 |           |

|    |    |      |     |    |        |                        |          |
|----|----|------|-----|----|--------|------------------------|----------|
| 21 | 46 | 70.8 | 424 | 9  | 047628 | TIER PROTEIN.          | 1.96e+01 |
| 22 | 46 | 70.8 | 513 | 10 | P97324 | GLUCOSE-6-PHOSPHATE DE | 1.96e+01 |
| 23 | 46 | 70.8 | 689 | 12 | 013265 | TISSUE TRANSGLUTAMINAS | 1.96e+01 |
| 24 | 45 | 69.2 | 176 | 3  | P90669 | CARBOXYPEPTIDASE-RELAT | 3.16e+01 |
| 25 | 45 | 69.2 | 194 | 4  | P79103 | RIBOSOMAL PROTEIN S4 ( | 3.16e+01 |
| 26 | 45 | 69.2 | 289 | 3  | P91237 | COSMID F08D12.         | 3.16e+01 |
| 27 | 45 | 69.2 | 346 | 3  | 002240 | HYPOTHEICAL PROTEIN F  | 3.16e+01 |
| 28 | 45 | 69.2 | 361 | 8  | 004009 | S-ADENOSYLMETHIONINE D | 3.16e+01 |
| 29 | 45 | 69.2 | 362 | 8  | 096555 | L-THREONINE DEAMINASE. | 3.16e+01 |
| 30 | 45 | 69.2 | 508 | 9  | P73375 | PLASMD DNA FOR HORA.   | 3.16e+01 |
| 31 | 45 | 69.2 | 583 | 9  | 032748 | VP2 GENOMIC RNA, COMPL | 3.16e+01 |
| 32 | 45 | 69.2 | 882 | 11 | 089813 | CASPID PROTEIN VP2.    | 3.16e+01 |
| 33 | 45 | 69.2 | 882 | 11 | 086218 | RNA BINDING PROTEIN.   | 3.16e+01 |
| 34 | 45 | 69.2 | 124 | 10 | 035335 | RNA BINDING PROTEIN.   | 3.16e+01 |
| 35 | 44 | 67.7 | 184 | 3  | 021775 | R06C7.6.               | 5.06e+01 |
| 36 | 44 | 67.7 | 189 | 2  | 014797 | HRS (FRAGMENT).        | 5.06e+01 |
| 37 | 44 | 67.7 | 203 | 9  | 030343 | HEMAGGLUTININ/PROTEASE | 5.06e+01 |
| 38 | 44 | 67.7 | 270 | 10 | 056167 | HRS.                   | 5.06e+01 |
| 39 | 44 | 67.7 | 310 | 9  | 056167 | BITUNCTIONAL CYCLASE/D | 5.06e+01 |
| 40 | 44 | 67.7 | 322 | 3  | 023796 | HNRP PROTEIN.          | 5.06e+01 |
| 41 | 44 | 67.7 | 344 | 2  | 013247 | SR55-1 PRE-MRNA SPLIC  | 5.06e+01 |
| 42 | 44 | 67.7 | 405 | 3  | 016735 | K09F6.8 PROTEIN.       | 5.06e+01 |
| 43 | 44 | 67.7 | 510 | 3  | 025168 | CYCLOIN B3.            | 5.06e+01 |
| 44 | 44 | 67.7 | 514 | 8  | 043727 | GLUCOSE-6-PHOSPHATE 1- | 5.06e+01 |
| 45 | 44 | 67.7 | 918 | 2  | 014159 | KIAA0146 PROTEIN (FRAG | 5.06e+01 |

## ALIGNMENTS

RESULT 1  
ID 029274; PRELIMINARY; PRT; 347 AA.  
AC 029274;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE IMMUNOGENIC PROTEIN (BCSP31-3).  
GN A0988.  
OS ARCHAEOGLOBUS FULGIDUS.  
OC ARCHAEOBACTERIA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KLEINK H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E.,  
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKY E.K., PETERSON J.D.,  
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KRAPIDES N.C.,  
RA KIRKNESS E.F., DOUGHERTY J., LEE N.H., SUTTON G.G., GILL S.,  
RA PETERSON S., REICH C.I., MCNEIL L.R., BADGER J.H., GLOER A., ZHOU L.,  
RA COTTON M.D., SPRIGGS T., ARTACH P., KAINE B.P., STYKS S.M.,  
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,  
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOSESE C.R.,  
RA VENTER J.C.,  
RA SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA KLEINK H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E.,  
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKY E.K., PETERSON J.D.,  
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KRAPIDES N.C.,  
RA KIRKNESS E.F., DOUGHERTY J., LEE N.H., SUTTON G.G., GILL S.,  
RA PETERSON S., REICH C.I., MCNEIL L.R., BADGER J.H., GLOER A., ZHOU L.,  
RA COTTON M.D., SPRIGGS T., ARTACH P., KAINE B.P., STYKS S.M.,  
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,  
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOSESE C.R.,  
RA VENTER J.C.,  
RA SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AEO01036; G2649610;  
SQ SEQUENCE 347 AA; 38112 MM; 764267E CRC32;  
Query Match 78.5%; Score 51; DB 9; Length 347;  
Best Local Similarity 75.0%; Pred. No. 1.58e+00; Mismatches 0; Gaps 0;  
Matches 6; Conservative 2; Indels 0;

Db 273 PEDAVYKL 280  
QY 2 PDDAVYKL 9

RESULT 2  
ID 099764 PRELIMINARY; PRT: 437 AA.

AC 099764;  
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
DT 01-NOV-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE SEMAPHORIN D (SEMAPHORIN III) (FRAGMENT).

OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.

RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN;

RX MEDLINE: 96207227.

RA ANDERSSON B., WENTLAND M.A., RICAFFENTE J.Y., LIU W., GIBBS R.A.;  
ANAL. BIOCHEM. 236:107-113(1996).

RL [2]

RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN;

RA YU W., ANDERSSON B., MORLEY K.C., MUZY D.M., DING Y., LIU W.,  
RICAFFENTE J.Y., WENTLAND M.A., LENNON G., GIBBS R.A.;

RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: U79241: G1710188;

KW HYPOTHETICAL PROTEIN.

FT NON\_TER 1

SO SEQUENCE 437 AA; 48483 MW; 4590C34B CRC32;

Query Match 76.9%; Score 51; DB 2; Length 437;  
Best Local Similarity 66.7%; Pred. No. 1.58e+00;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 211 QASEAVYKL 219

QY 1 QPDDAVYKL 9

RESULT 3  
ID 062215 PRELIMINARY; PRT: 666 AA.

AC 062215;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE SEMAPHORIN D (SEMAPHORIN III) (FRAGMENT).

GN SEMAD OR SEMA III.

OS MUS MUSCULUS (MOUSE).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; RODENTIA.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 95267432.

RA MESSERSMITH E.K., LEONARDO E.D., SHATZ C.J., TESSIER-LAVIGNE M.,  
GOODMAN C.S., KOLODKIN A.L.;

RL NEURON 14:949-959(1995).

DR EMBL: L40484: G703190;

DR MGD: MGI:107558; SEMAD.

FT NON\_TER 1

SO SEQUENCE 666 AA; 76452 MW; E441E282 CRC32;

Query Match 76.9%; Score 50; DB 10; Length 666;  
Best Local Similarity 44.4%; Pred. No. 2.66e+00;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 43 HPEDNIFKL 51

QY 1 OPDDAVYKL 9

RESULT 4  
ID 014563 PRELIMINARY; PRT: 771 AA.

AC 014563;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)

DE SEMAPHORIN-III.

GN SEMA-III.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN;

RX MEDLINE: 94094332.

RA KOLODKIN A.L., MATTHES D.J., GOODMAN C.S.;

RL CELL 75:1389-1399(1993).

DR EMBL: L26081: G436560;

SO SEQUENCE 771 AA; 88889 MW; 9EB1A137 CRC32;

Query Match 76.9%; Score 50; DB 2; Length 771;  
Best Local Similarity 44.4%; Pred. No. 2.66e+00;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 149 HPEDNIFKL 157

QY 1 QPDDAVYKL 9

RESULT 5  
ID 008665 PRELIMINARY; PRT: 772 AA.

AC 008665;

DT 01-JUL-1997 (TREMBLREL. 04, CREATED)

DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)

DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)

DE COLLAPSED-1.

OS MUS MUSCULUS (MOUSE).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; RODENTIA.

RN [1]

RP SEQUENCE FROM N.A.

RC TANGUCHI M.;

RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: D85028; D1020554;

SO SEQUENCE 772 AA; 88799 MW; 4F0698CF CRC32;

Query Match 76.9%; Score 50; DB 10; Length 772;  
Best Local Similarity 44.4%; Pred. No. 2.66e+00;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 149 HPEDNIFKL 157

QY 1 QPDDAVYKL 9

RESULT 6  
ID 063548 PRELIMINARY; PRT: 772 AA.

AC 063548;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)

DE SEMAPHORIN III/COLLAPSED-1.

OS RATTUS NORVEGICUS (RAT).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; RODENTIA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-MISTAR; TISSUE-BRAIN;

RA GIGER R.I.;

RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: X95286; E220165;

SO SEQUENCE 772 AA; 88808 MW; 7E8BFD35 CRC32;

```

RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SAOM P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.,
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AE001103; G2650604; -.
KW TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ SEQUENCE 1213 AA; 140592 MW; B63AD3D1F CRC32;

Query Match 76.9%; Score 50; DB 9; Length 1213;
Best Local Similarity 62.5%; Pred. No. 2,66e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 730 PNEVYKRL 737
QY 2 PDDAVYKRL 9
 1:::1111

RESULT 9
ID P74619 PRELIMINARY; PRT; 832 AA.
AC P74619;
DT 01-FEB-1997 (TREMBLEREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLEREL. 02, LAST SEQUENCE UPDATE)
DE 01-FEB-1997 (TREMBLEREL. 02, LAST ANNOTATION UPDATE)
DE SYNECHOCYSTIS SP.
OS SYNECHOCYSTIS SP.
OC EUBACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.,
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROKAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSODUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NAKUO K.,
RA OKUMURA S., SHIMO S., TAKEUCHI C., WADA T., WATANABE A.,
RA YASUDA M., YASUDA M., TABATA S.,
RL DNA RES. 3:109-136(1996).
DR EMBL; D90916; G1653816; -.
KW HYPOTHEICAL PROTEIN.
SQ SEQUENCE 832 AA; 92865 MW; 08554B2A CRC32;

Query Match 75.4%; Score 49; DB 9; Length 832;
Best Local Similarity 44.4%; Pred. No. 4,43e+00;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 309 QPDEQIFRL 317
QY 1 QPDDAVYKRL 9
 1111:::1111

RESULT 10
ID Q22670 PRELIMINARY; PRT; 1307 AA.
AC Q22670;
DT 01-NOV-1996 (TREMBLEREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLEREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLEREL. 01, LAST ANNOTATION UPDATE)
DE T22C1.10.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RL MCOURRAY A.,
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
AC MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BENFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,
RA COULSON A., CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,

```

RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
 RA LATREILLE P., LIGHNING J., LLOYD C., MCGRATH A., MORTIMORE B.,  
 RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFEEN L., ROOPRA A.,  
 RA SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONNHAMMER E.,  
 RA STADEN R., SULTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
 RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
 RA WIKINSON-SPROAT J., WOHLDMAN P.,  
 RL NATURE 368:32-38(1994).  
 DR EMBL: Z75350; E250299;  
 SQ SEQUENCE 1307 AA; 149908 MW; 5451AC24 CRC32;

Query Match 72.3%; Score 48; DB 3; Length 1307;  
 Best Local Similarity 62.5%; Pred. No. 7.32e+00;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 118 PDEPVYKI 125  
 111:111:  
 QY 2 PDDAVYKL 9

RESULT 11  
 ID 064288 PRELIMINARY; PRT; 163 AA.  
 AC 064288;  
 DT 01-JAN-1998 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE OLFACTOR MARKER PROTEIN.  
 GN OMP.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94154378.  
 RA BROWN K.A., SUTCLIFFE M.J., STEELE K., BROWN S.D.;  
 RL MAMM. GENOME 5:11-14(1994).  
 [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-129;  
 RX MEDLINE: 94307732.  
 RA BODIAKOVA O.I., RAMA KRISHNA N., GETCHELL T.V., MARGOLIS F.L.;  
 RL GENOMICS 20:452-462(1994).  
 DR EMBL: U02557; G493517;  
 DR EMBL: U01213; G520741;  
 DR MGD: MGI:97436; OMP.  
 SQ SEQUENCE 163 AA; 18866 MW; 4ECB75E0 CRC32;

Query Match 72.3%; Score 47; DB 10; Length 163;  
 Best Local Similarity 44.4%; Pred. No. 1.20e+01;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 47 RPAESVYRL 55  
 111:111:  
 QY 1 QPDDAVYKL 9

RESULT 12  
 ID 024358 PRELIMINARY; PRT; 317 AA.  
 AC 024358;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD) (FRAGMENT).  
 GN G6PD.  
 OS SPINACIA OLERACEA (SPINACH).  
 OC PLASMID PZL1.  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;  
 OC CAROPHYLLALES; CHENOPODIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LEAVES; STRAIN-CV. MATADOR;  
 RA FLINK A., DIOGON T., PERROUD P.F., CRESPI P., GREPPIN H.;

RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) -  
 CC D-GLUCONO-DELTA-LACTONE 6-PHOSPHATE + NADPH.  
 CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.  
 DR EMBL: AJ000183; E330370;  
 DR PROSITE: P500069; G6P-DEHYDROGENASE; 1.  
 KW OXIDOREDUCTASE; PLASMID; NADP; GLUCOSE METABOLISM.  
 FT NON\_TER 1  
 FT ACT\_SITE 7  
 SQ SEQUENCE 317 AA; 36049 MW; AFB86F6 CRC32;

Query Match 72.3%; Score 47; DB 8; Length 317;  
 Best Local Similarity 71.4%; Pred. No. 1.20e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 202 QPDEAVY 208  
 111:111:  
 QY 1 QPDDAVY 7

RESULT 13  
 ID 024359 PRELIMINARY; PRT; 465 AA.  
 AC 024359;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD) (FRAGMENT).  
 GN G6PD.  
 OS SPINACIA OLERACEA (SPINACH).  
 OC PLASMID PZL1.  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;  
 OC CAROPHYLLALES; CHENOPODIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LEAVES; STRAIN-CV. MATADOR;  
 RA FLINK A., DIOGON T., PERROUD P.F., CRESPI P., GREPPIN H.;  
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) -  
 CC D-GLUCONO-DELTA-LACTONE 6-PHOSPHATE + NADPH.  
 CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.  
 DR EMBL: AJ000184; E330368;  
 DR PROSITE: P500069; G6P-DEHYDROGENASE; 1.  
 KW OXIDOREDUCTASE; PLASMID; NADP; GLUCOSE METABOLISM.  
 FT NON\_TER 1  
 FT ACT\_SITE 7  
 FT ACT\_SITE 225  
 SQ SEQUENCE 465 AA; 53101 MW; 929DA6AC CRC32;

Query Match 72.3%; Score 47; DB 8; Length 465;  
 Best Local Similarity 71.4%; Pred. No. 1.20e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 420 QPDEAVY 426  
 111:111:  
 QY 1 QPDDAVY 7

RESULT 14  
 ID 043728 PRELIMINARY; PRT; 492 AA.  
 AC 043728;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD) (FRAGMENT).  
 GN G6PDH.  
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;  
 OC CARPALES; CRUCIFERAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA; TISSUE-LEAVES, STEMS, SOME FLOWERS AND ROOTS;  
 RA FLINK A., GREPPIN H., TACCHINI P.;  
 RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) =  
 CC D-GLUCONO-DELTA-LACTONE 6-PHOSPHATE + NADPH.  
 CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.  
 DR EMBL: X84229; E137419;  
 DR PROSITE: PS00069; G6P\_DEHYDROGENASE; 1.  
 KW OXIDOREDUCTASE; NADP; GLUCOSE METABOLISM.  
 FT NON\_TER 1  
 FT ACT\_SITE 183  
 SQ SEQUENCE 492 AA; 56203 MW; 8D214E1E CRC32;

Query Match 72.3%; Score 47; DB 8; Length 492;  
 Best Local Similarity 71.4%; Pred. No. 1.20e+01;  
 Matches: 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB: 378 OPDDAVY 384  
 11:1:1  
 QY 1 OPDDAVY 7



















RESULT 15  
 ID: 027464 PRELIMINARY: PRT; 522 AA.  
 AC: 027464;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD).  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA WHITE S.;  
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,  
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STILSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
 RL NATURE 368:32-38(1994).  
 CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) =  
 CC D-GLUCONO-DELTA-LACTONE 6-PHOSPHATE + NADPH.  
 CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.  
 DR EMBL: Z73102; E242588;  
 DR PROSITE: PS00069; G6P\_DEHYDROGENASE; 1.  
 KW OXIDOREDUCTASE; NADP; GLUCOSE METABOLISM.  
 FT ACT\_SITE 211  
 FT BY SIMILARITY.  
 SQ SEQUENCE 522 AA; 60215 MW; D82B2BCE CRC32;

Query Match 72.3%; Score 47; DB 3; Length 522;  
 Best Local Similarity 62.5%; Pred. No. 1.20e+01;  
 Matches: 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

DB: 403 OPENAYYM 410  
 11:1:1  
 QY 1 OPDDAVY 8

Search completed: Fri Sep 11 13:04:02 1998  
 Job time: 37 secs.

**THIS PAGE BLANK (USPTO)**

Release 3.1A John F. Collins, Biocomputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Fri Sep 11 12:57:43 1998;  MasPar time 2.57 Seconds
```

Tabular output not generated.

|                |                           |
|----------------|---------------------------|
| Title:         | >US-08-452-843-6          |
| Description:   | (1-9) from US08452843.pdf |
| Document Name: | 75                        |

Sequence: 1 FAMPNFYTL 9

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq32

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 17.002; Variance 57.332; scale 0.297

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description              | Pred. No. |
|------------|-------|-------------|--------|----|--------|--------------------------|-----------|
| 1          | 75    | 100.0       | 9      | 18 | R89367 | Cw3 consensus peptide    | 3.80e-01  |
| 2          | 57    | 76.0        | 9      | 18 | R89366 | Cw3 consensus peptide    | 2.78e-01  |
| 3          | 54    | 72.0        | 1103   | 8  | R39631 | Neurofibromatosis type 1 | 5.52e-01  |
| 4          | 54    | 72.0        | 2485   | 11 | R59921 | RAS associated GAP NF    | 5.52e-01  |
| 5          | 54    | 72.0        | 2485   | 11 | R59922 | RAS associated GAP NF    | 5.52e-01  |
| 6          | 54    | 72.0        | 2818   | 21 | M13380 | Human neurofibromin      | 5.52e-01  |
| 7          | 54    | 72.0        | 2818   | 21 | R23258 | NF1 gene product         | 5.52e-01  |
| 8          | 53    | 70.7        | 741    | 15 | R89327 | Membrane anchor prote    | 6.82e-01  |
| 9          | 51    | 68.0        | 1464   | 17 | R88469 | Feline infectious per    | 1.08e-02  |
| 10         | 50    | 66.7        | 386    | 29 | W5311  | H. pylori ORF 05epi03    | 1.35e-02  |
| 11         | 50    | 66.7        | 884    | 17 | R93021 | Human glucagon degraded  | 1.35e-02  |
| 12         | 49    | 65.3        | 409    | 8  | R41227 | 910 SLG protein          | 1.69e-02  |
| 13         | 49    | 65.3        | 858    | 10 | R53404 | S-Locus receptor (ser    | 1.69e-02  |
| 14         | 49    | 65.3        | 1399   | 8  | R38698 | SPRV-05 TGE virus g      | 2.10e-02  |
| 15         | 48    | 64.0        | 3011   | 16 | R95021 | Hepatitis GB virus G     | 2.10e-02  |
| 16         | 47    | 62.7        | 38     | 9  | R46636 | 70 kD proteoglycan co    | 2.62e-02  |
| 17         | 47    | 62.7        | 38     | 9  | R46637 | 70 kD proteoglycan co    | 2.62e-02  |
| 18         | 47    | 62.7        | 219    | 10 | R51799 | Chloramphenicol-acye     | 2.62e-02  |

|    |    |      |      |    |        |                                   |          |
|----|----|------|------|----|--------|-----------------------------------|----------|
| 19 | 47 | 62.7 | 219  | 10 | R51278 | Chloramphenicol- $\alpha$ -acetyl | 2.52e+02 |
| 20 | 47 | 62.7 | 240  | 1  | R05425 | Amino acid sequence f             | 2.52e+02 |
| 21 | 47 | 62.7 | 241  | 1  | P93070 | Sequence of chloramph             | 2.52e+02 |
| 22 | 47 | 62.7 | 249  | 29 | P93358 | Apoptosis associated              | 2.52e+02 |
| 23 | 47 | 62.7 | 230  | 1  | P92068 | Fusion protein associat           | 2.52e+02 |
| 24 | 47 | 62.7 | 231  | 1  | R16511 | SP-C from CP210SP-C 1             | 2.52e+02 |
| 25 | 47 | 62.7 | 231  | 1  | R05419 | CAR:SP-C hybrid prote             | 2.52e+02 |
| 26 | 47 | 62.7 | 293  | 1  | P05418 | CAR:SP-B hybrid prote             | 2.52e+02 |
| 27 | 47 | 62.7 | 402  | 1  | P81179 | Sequence of human end             | 2.52e+02 |
| 28 | 47 | 62.7 | 412  | 11 | R36447 | TMV replicon-encod                | 2.52e+02 |
| 29 | 47 | 62.7 | 930  | 23 | R18061 | Pasteurella haemolyti             | 2.52e+02 |
| 30 | 46 | 61.3 | 319  | 11 | R56870 | Catine gastric lipase             | 3.25e+02 |
| 31 | 46 | 61.3 | 379  | 23 | M09382 | Dog gastric lipase pr             | 3.25e+02 |
| 32 | 46 | 61.3 | 380  | 11 | R56871 | Canine gastric lipase             | 3.25e+02 |
| 33 | 46 | 61.3 | 526  | 18 | R95697 | Erythrobacter longus              | 3.25e+02 |
| 34 | 46 | 61.3 | 562  | 16 | R86407 | Human matrix metallo              | 3.25e+02 |
| 35 | 46 | 61.3 | 582  | 14 | R75648 | Human placenta derive             | 3.25e+02 |
| 36 | 46 | 61.3 | 594  | 22 | W20603 | H. pylori secreted or             | 3.25e+02 |
| 37 | 46 | 61.3 | 919  | 23 | P18580 | Potato alpha-glucosid             | 3.25e+02 |
| 38 | 46 | 61.3 | 925  | 14 | R79148 | Human insulin recepto             | 3.25e+02 |
| 39 | 46 | 61.3 | 985  | 8  | R42214 | Aspergillus niger glu             | 3.25e+02 |
| 40 | 46 | 61.3 | 985  | 8  | R42995 | Glycosyltransferase.              | 3.25e+02 |
| 41 | 46 | 61.3 | 985  | 22 | W15191 | Aspergillus oryzae al             | 3.25e+02 |
| 42 | 46 | 61.3 | 1027 | 22 | W10757 | Candida albicans chit             | 4.03e+02 |
| 43 | 45 | 60.0 | 286  | 29 | W46265 | Sensory and motor neu             | 4.03e+02 |
| 44 | 45 | 60.0 | 286  | 28 | W41263 | Sensory and motor neu             | 4.03e+02 |
| 45 | 45 | 60.0 | 379  | 27 | W26714 | Plasminogen activator             | 4.03e+02 |

## ALIGNMENTS

RESULT 1  
ID R89367 standard; peptide: 9 AA.  
AC R89367;  
DT 18-SEP-1996 (first entry)  
DE Cys consensus peptide derived immunogenic peptide #2.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN MO9603140-A1.  
PN 08-FEB-1996.  
PF 21-JUL-1995; 009234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
PT WPI: 96-116784/12.  
DR Compr. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PR and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2: Page 26: 32pep: English  
CC The sequences given in R89367-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g. the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

|                       |         |                     |        |               |
|-----------------------|---------|---------------------|--------|---------------|
| Query Match           | 100.0%; | Score 75;           | DB 18; | Length 9;     |
| Best Local Similarity | 100.0%; | Pred. No. 3.80e-01; |        |               |
| Matches               | 9;      | Conservative        | 0;     | Mismatches 0; |
|                       |         |                     |        | Indels 0;     |
|                       |         |                     |        | Gaps 0;       |

|    |   |           |   |
|----|---|-----------|---|
| Db | 1 | fampnfytl | 9 |
|    |   |           |   |
| QY | 1 | FAMPNFTTL | 9 |

|                       |              |                     |               |              |
|-----------------------|--------------|---------------------|---------------|--------------|
| Query Match           | 72.0%        | Score 54;           | DB: 8;        | Length 1103; |
| Best Local Similarity | 55.6%        | Pred. No. 5.52e+01; |               |              |
| Matches 5;            | Conservative | 3;                  | Mismatches 1; | Indels 0;    |
|                       |              |                     | Gaps 0;       |              |
| Db                    | 551          | ts1pkfy11           | 559           |              |
|                       |              |                     |               |              |

|    |   |                            |  |
|----|---|----------------------------|--|
| AC | DT  | 5                          | RESULT   |
| AC | DT  | 22-FEB-1995                | (first entry)                                      |
| DE | RAS   | associated GAP             | NF204.   |
| KW | pkrt10  | gpkrase activating protein | GAP; GAP related domain; GRD;                      |
| KW | pkrt10  | pkpkl                      | Saccharomyces cerevisiae; RAS2; v-Ras; heat shock; |
| KW | neurofibromatosis   | type 1; NFI.               |  |
| OS | Homo sapiens.   |                            |  |
| PN | MO9416069-A.  |                            |  |
| PE | 21-JUL-1994.  |                            |  |
| PE | 12-JAN-1994   | U00198.                    |  |
| PR | 15-JAN-1993   | US-004824.                 |  |
| PA | (SCHE )   | SCHERING CORP.             |  |
| PI | Kaziro Y,   | Nakafuku M;                |  |
| DR | NFI   | 94-249216/30.              |  |
| PT | Blocking Ras-induced effects on a cell - by introducing a gpkrase   |                            |  |
| PT | activating protein to the cell, used esp. in treatment of cancers   |                            |  |
| PS | disclosure  | Page 44-52; 87pp           | English.   |
| CC | Human neurofibromatosis type 1 (NFI)-GAP related domain (GRD) was   |                            |  |
| CC | cloned into the yeast expression vector pkrt10 to obtain pkpkl.     |                            |  |
| CC | pkpkl DNA was mutagenized by hydroxylamine in vitro and transformed |                            |  |
| CC | into S. cerevisiae tk161-R2V-D, which carries an oncogenic-type     |                            |  |
| CC | RAS2Val19 mutation. The heat shock sensitivity of the clones was    |                            |  |
| CC | checked. Plasmid DNAs were recovered, re-transformed into tk161-    |                            |  |
| CC | (R2V-D), and phenotypic reversal was examined. 2 clones, NF201      |                            |  |
| CC | and NF204, which had strong suppression activity                    |                            |  |
| CC | for RAS2Val19, were selected. The mutant NFI-GRDs were also able    |                            |  |
| CC | to inhibit v-Ras-induced transformation in mammalian cells.         |                            |  |
| CC | Sequence  | 2485 AA;                   |  |
| CC |   |                            |  |



Query Match 72.0%; Score 54; DB 11; Length 2485;

Best Local Similarity 55.6%; Pred. No. 5.52e+01; Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 1802 fsipkfy11 1810

QY 1 FAMPNRYTL 9

RESULT 6  
M13280 standard; Protein: 2818 AA.

AC M13280;  
DT 05-JUN-1997 (first entry)

DE Human neurofibromin.  
KW Human; neurofibromin; yeast; IRA; protein; inhibition; GTPase;  
KW regulation; ras-cAMP; pathway; mammalian; GAP; ras p21; gene;  
KW activation; neurofibromatosis; type 1; NFI; somatic; mutation;  
KW tumour; detection; diagnosis; prognosis; defective; treatment;  
OS Homo sapiens.

PH Key Location/Qualifiers  
FT domain 1175..1534  
FT /note= "GTPase activating protein (GAP) related  
FT domain (GRD)"  
FT 1389..1391  
FT /note= "conserved region in GRD"

region  
FT US5605799-A.  
PD 25-FEB-1997.  
PF 12-JUL-1990; 551531.  
PR 12-JUL-1990; US-551531.  
PR 16-APR-1993; US-047088.  
PR 28-MAR-1995; US-411389.  
PA (UTAH ) UNIV UTAH RES FOUND.  
PI Cawthon RM, Li Y, White RL.  
DR WPI: 97-153572/14.  
DR N-PSDB: T46941.

PT Detection of defective ras regulation at the neurofibromatosis type  
PT 1 gene in tumour by detecting mutation in specified region of gene  
PS Claim 1; Columns 17-38; 35pp; English.

CC The present sequence is human neurofibromin (hNF), which is  
CC largely homologous to yeast IRA protein (inhibitory regulators of  
CC the ras-cAMP pathway) and mammalian GAP (ras p21 GTPase activating  
CC proteins). The hNF gene is the human neurofibromatosis type 1 (NFI)  
CC gene, somatic mutations of which in the region spanning nucleotides  
CC 3803-4868 of the NFI cDNA, in human tumours, indicates defective  
CC mutation in the NFI gene can be treated using ras activity as the  
CC focus, whereas a tumour not containing such a mutation will require  
CC other courses of treatment. A tumour containing a somatic mutation  
CC in the NFI gene can be treated by inactivating ras p21, also as GAP  
CC p120 is present, but apparently latent, GAP p120 activation would  
CC be beneficial and finally inhibition of GDP/GTP exchange would also  
CC counteract the loss of hNF or hNF GAP related domain activity.  
SQ Sequence 2818 AA;

Query Match 72.0%; Score 54; DB 21; Length 2818;

Best Local Similarity 55.6%; Pred. No. 5.52e+01; Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 2135 fsipkfy11 2143

QY 1 FAMPNRYTL 9

RESULT 7  
R22268 standard; Protein: 2818 AA.

AC R22268;  
DT 06-MAY-1992 (first entry)

DE NFI gene product.  
KW von Recklinghausen neurofibromatosis disease; autosomal dominant;  
KW gene therapy.  
OS Homo sapiens.  
PN MO9200387-A.

PD 09-JAN-1992.  
PF 28-JUN-1993; U04624.  
PR 29-JUN-1990; US-547090.

PA (UNMI ) UNIV OF MICHIGAN.  
PI Collins FS, Wallace MR, Marchuk DA, Andersen LB, Gutmann DH;  
DR WPI: 92-041568/05.  
DR N-PSDB: 020602.

PT DNA sequences to von-Recklinghausen neurofibromatosis gene - and  
PT derived amino acid sequences and probes for screening NFI in early  
PT stages of disease  
PS Claim 25; Page 67; 122pp; English.

CC This is the amino acid sequence of the von Recklinghausen neuro-  
CC fibromatosis (NFI) gene product. It and antibodies raised to it  
CC can be used in hybridisation and immunological assays to screen for  
CC the presence of a normal or defective NFI gene product. Functional  
CC assays to measure levels of gene function can also be used for  
CC diagnosis or to monitor treatment. Patient therapy through  
CC supplementation with the normal NFI product which can be  
CC produced by recombinant techniques is also possible.  
SQ Sequence 2818 AA;

Query Match 72.0%; Score 54; DB 4; Length 2818;

Best Local Similarity 55.6%; Pred. No. 5.52e+01; Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 2135 fsipkfy11 2143

QY 1 FAMPNRYTL 9

RESULT 8  
R89327 standard; Protein: 741 AA.

AC R89327;  
DT 08-APR-1996 (first entry)

DE Membrane anchor protein. Integrase; hpc gene; p11n protein;  
KW LKPI operon; peptidase; periplasmic chaperone protein;  
KW minor tip-associated protein; tip adhesin protein; cloning;  
KW Escherichia coli; plasmid pHFI; diagnostic; probe; antibody;  
KW recombinant vaccine.  
OS Haemophilus influenzae (serotype 1).

PN MO9602648-A1.

PD 01-FEB-1996.  
PF 13-JUL-1995; U08789.  
PR 19-JUL-1994; US-277231.

PR 07-JUN-1995; US-477326.  
PR 07-JUN-1995; US-473750.

PA (AMCY ) AMERICAN CYANAMID CO.  
PI (BACT-) BACTEX INC.  
PI Brinton CC, Green BA;

DR WPI: 96-105910/11.  
DR N-PSDB: 099312.

PT Haemophilus influenzae 1 LKPI p11n genes and proteins - used to  
PT produce anti-H. influenzae antibodies, used to detect and vaccinate  
PT against H. influenzae  
PS Claim 2; Page 45-47; 63pp; English.

CC The sequence represents a membrane anchor protein encoded by the  
CC hpc gene in the LKPI operon from Haemophilus influenzae serotype-1.  
CC The operon also encodes integrase, p11n protein (R89325),  
CC periplasmic chaperone protein (R89326), minor tip-associated protein  
CC (R89328), tip adhesin protein (R89329) and peptidase. The operon  
CC has been isolated by cloning in Escherichia coli using plasmid pHFI.  
CC The operon and its encoded proteins may be used in production of  
CC diagnostic probes, antibodies and recombinant vaccines.  
SQ Sequence 741 AA;

Query Match 70.7%; Score 53; DB 15; Length 741;

Best Local Similarity 55.6%; Pred. No. 6.92e+01; Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 368 ysspdfy11 376

QY 1 FAMPNRYTL 9

RESULT 9  
 ID R88469 standard: Protein; 1464 AA.  
 AC R88469.  
 DT 14-AUG-1996 (first entry)  
 DE Feline infectious peritonitis 1 virus spike protein.  
 KW Feline infectious peritonitis 1 virus; FIPV-I; spike protein;  
 KM vaccine; prevention; treatment.  
 OS Feline infectious peritonitis 1 virus.  
 PN J0732/683-A.  
 PD 19-DEC-1995. 129300.  
 PE 10-JUN-1994; JP-129300.  
 PR 10-JUN-1994; JP-129300.  
 PA (KITA) KITASATO KENKYUSHO SH.  
 WP: 96-072341/08.  
 DR N-PSDB: T10106.  
 PT DNA encoding feline infectious peritonitis 1 virus spike protein -  
 used in a vaccine for prevention and treatment of FIPV-I infection  
 PS Claim 1; Page 14-17, 23pp; Japanese.  
 CC This sequence represents the feline infectious peritonitis 1 virus  
 (FIPV-I) spike protein. The FIPV-I spike protein may be used in the  
 CC production of a vaccine for the prevention and treatment of FIPV-I  
 CC infection. The spike protein may be produced by transforming a host  
 CC cell with the spike protein DNA and expressing the sequence such  
 CC that the spike protein can be isolated.  
 SQ Sequence 1464 AA.

Query Match 68.0%; Score 51; DB 17; Length 1464;  
 Best Local Similarity 57.1%; Pred. No. 1.08e+02;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 800 ytmphy 806  
 1 FAMPHY 7

RESULT 10  
 ID W55311 standard: Protein; 386 AA.  
 AC W55311.  
 DT 15-JUN-1998 (first entry)  
 DE H. pylori ORF 06epi0306orf11 protein.  
 KW Cytoplasmic; vaccine; prevention; infection; envelope;  
 KM identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
 OS Helicobacter pylori.  
 PN W0973704-A1.  
 PD 09-OCT-1997.  
 PE 27-MAR-1997; U05223.  
 PR 06-DEC-1996; US-761318.  
 PR 29-MAR-1996; US-625811.  
 PR 02-APR-1996; US-758731.  
 PR 23-OCT-1996; US-736905.  
 PR 28-OCT-1996; US-738859.  
 PA (ASTR) ASTRA AB.  
 PI Alm RA, Smith D;  
 WP: 97-503122/46.  
 DR N-PSDB: V24720.  
 PT Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 PS Claim 14; Page 540-541; 1145pp; English.  
 CC This sequence is a H. pylori protein of unspecified function.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated

CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions for  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.

Sequence 386 AA;

Query Match 66.7%; Score 50; DB 29; Length 386;  
 Best Local Similarity 57.1%; Pred. No. 1.35e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 164 famphy 170  
 1 FAMPHY 7

RESULT 11  
 ID R93021 standard: Protein; 864 AA.  
 AC R93021.  
 DT 09-AUG-1996 (first entry)  
 DE Human glucagon degrading enzyme.  
 KW Glucagon degrading enzyme; catalyst; cleavage; selectin; human; primer;  
 KW vasoactive intestinal peptide; VIP; pancreatic carcinoma cell line; PCR;  
 KW amplification; polymerase chain reaction; probe; expression vector;  
 KW eukaryote; SV40 promoter; COS-7.  
 OS Homo sapiens.  
 PN J08023972-A.  
 PD 30-JAN-1996.  
 PE 19-JUL-1994; 187936.  
 PR 19-JUL-1994; JP-187936.  
 PA (SUNR) SUNTORY LTD.  
 WP: 96-133414/14.  
 DR N-PSDB: T11575.  
 PT New glucagon decomposing enzyme, and DNA encoding it - for  
 PT specifically cleaving glucagon and vasoactive intestinal peptide, in  
 PT the prevention and treatment of diseases caused by excess glucagon  
 PT and VIP  
 PS Claim 2; Page 2, 18pp; Japanese.  
 CC This is the amino acid sequence of a novel isolated glucagon degrading  
 CC enzyme (GDE) of mol. wt. 83 KD. The enzyme has a pH optimum of 6.8 and  
 CC catalyses the cleavage of glucagon, vasoactive intestinal peptide and  
 CC selectin (R93022-4). The corresp. gene was isolated from a human  
 CC pancreatic carcinoma cell line HPC-YO cDNA library by screening the  
 CC library with an anti-GDE peptide antibody, amplifying the inserts with  
 CC the primers T18903-4 and probing the fragments with the probe T18905.  
 CC This screening resulted in the full length clone designated lambda  
 CC GDE4-2. The coding region of the clone was subsequently PCR amplified by  
 CC the primers T11576-7 and inserted into the eukaryotic expression vector  
 CC pXDCR under control of the SV40 promoter for production of the protein in  
 CC COS-7 cells. The protein is useful in preventing and treating diseases  
 CC characterised by an excess of glucagon or vasoactive intestinal peptide.  
 SQ Sequence 864 AA;

Query Match 66.7%; Score 50; DB 17; Length 864;  
 Best Local Similarity 55.6%; Pred. No. 1.35e+02;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 412 famphy 420

Page 5

|         |                |
|---------|----------------|
| protein | 32..856        |
| FT      | /label=        |
| FT      | mature SRK-910 |
| FT      |                |

| FT | modified_site | 48.50 | /label= | potential N-glycosylation site |
|----|---------------|-------|---------|--------------------------------|
| FT |               |       |         |                                |
| FT |               |       |         |                                |

FT      /label= potential N-glycosylation site

|    |  |
|----|--|
| FT | /label= potential N-glycosylation site |
| FT | 152..154                               |
| FT | /label= potential N-glycosylation site |
| FT | 248..250                               |

FT /label= potential N-glycosylation site

|    |               |                                |
|----|---------------|--------------------------------|
| FT | modified_site | 318..320                       |
| FT | /label=       | potential N-glycosylation site |

|    |               |                                |
|----|---------------|--------------------------------|
| FT | modified_site | 393..395                       |
| FT | /label=       | potential N-glycosylation site |

| FT | region | 303 | /label= conserved residue |
|----|--------|-----|---------------------------|
| FT |        |     |                           |

| FT | region | 309 | /label= conserved residue |
|----|--------|-----|---------------------------|
| FT |        |     |                           |

| FT | region | 315 | /label= conserved residue |
|----|--------|-----|---------------------------|
| FT |        |     |                           |

| FT | region | 323 | /label- | conserved | residue |
|----|--------|-----|---------|-----------|---------|
| FT |        |     |         |           |         |

| FT | region | 325 | /label= conserved residue |
|----|--------|-----|---------------------------|
| FT |        |     |                           |

| FT | region | 346 | /label= conserved residue |
|----|--------|-----|---------------------------|
| FT |        |     |                           |
| FT |        |     |                           |

| FT | region | 354 | /label= conserved residue |
|----|--------|-----|---------------------------|
| FT |        |     |                           |

| FT | region | 384 | /label= conserved residue |
|----|--------|-----|---------------------------|
| FT |        |     |                           |

| FT | region | 388 | /label= conserved residue |
|----|--------|-----|---------------------------|
| FT |        |     |                           |

| FT | region | 392 | /label= conserved residue |
|----|--------|-----|---------------------------|
| FT |        |     |                           |

| FT | region | 394 | . | conserved residue |
|----|--------|-----|---|-------------------|
| FT |        |     |   | /label=           |

| FT | region | 409 | /label= conserved residue |
|----|--------|-----|---------------------------|
| FT |        |     |                           |

PN W09409139-A.  
PD 28-APR-1994.

PF 08-OCT-1993; U09448.  
PR 08-OCT-1992; US-959945.

PA (PION-) PIONEER HI-BRED INT INC.  
PA (UYGU-) UNIV GUELPH.

PI Goring D, Rothstein SJ;  
DR WPI; 94-151324/18.

DR N-PSDB; Q63492.  
PT DNA encoding the S-locus receptor kinase-910 protein -

PT screen for the self-incompatibility phenotype in plant  
PT confer the phenotype on self-compatible plants

PS Clalm 9; Flg 4A-4B; 72pp; English.  
CC R53404 encodes the S-locus receptor (serine/threonine)

cc (SRK910). This is associated with the self-incompatibility  
cc phenotype. SRK genes show homology with S-locus glycoproteins

cc The invention is useful for deriving probes that can be  
cc rapidly screen the progeny of cross fertilisations betw

CC species for the SI phenotype. The DNA can also be used  
CC the SI phenotype on a self-compatible plant.

Sequence 858 AA;

|                       |        |           |           |        |
|-----------------------|--------|-----------|-----------|--------|
| Query Match           | 65.3%; | Score 49; | DB 10;    | Length |
| Best Local Similarity | 62.5%; | Pred. No. | 1.69e+02; |        |

Matches 5; Conservative 2; Mismatches 1; Indels 0

Db 209 gmpetyl 216  
:|:|:|

QY 2 AMPNEYTL 9

```

RESULT 14
ID R36698 standard; Protein: 1399 AA.
AC R36698: 1399
DE 23-NOV-1993 (first entry)
S-PRV-055 TGE virus gp195 gene product.
KW Attenuated; vaccine; herpes virus; non-primate; live; safer; IBR;
infectious bovine rhinotracheitis; MDV; Marek's disease virus; fowl;
KW Pseudo-rabies; swine; transmissible gastroenteritis virus.
OS Transmissible gastroenteritis virus.
UN US523424.A
ED 29-JUN-1993
PE 27-JUL-1988; 225032.
PR 06-SEP-1985; US-773430.
PR 27-JAN-1986; US-823102.
PR 17-JUL-1986; US-887140.
PR 02-SEP-1986; US-902887.
PR 20-NOV-1986; US-933107.
PR 27-JUL-1987; US-078519.
PR 27-JUL-1988; US-225032.
PA (PRVT-) PRUTECH RES & DEV.
PI Chiang CH, Cochran MD, Macdonald RD;
DR WPI: 93-219585/27.
N-PSDB: Q42755.
PR Recombinant fusion proteins for vaccine - comprises antigenic
PT sequences fused to viral sequences e.g. pseudo-rabies virus, used
PT as vaccines
PS Disclosure: Fig 22; 127pp; English.
CC The sequence is that encoded by the transmissible gastroenteritis
CC virus gp195 gene from S-PRV-055 which may be used in the prodn.
CC of attenuated non-primate herpes viruses. These can be used as live
CC vaccines and provide a safer vaccine than currently available for
CC e.g. pseudorabies virus of swine, infectious bovine rhinotracheitis
CC (IBR) virus or Marek's disease of fowl.
SO Sequence 1399 AA;

Query Match 65.8%; Score 49; DB 8; Length 1399;
Best Local Similarity 71.4%; Pred. No. 1.69e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 510 fpmfyt1 516
OY 3 fpmfyt1 9

RESULT 15
ID R95021 standard; Protein: 3011 AA.
AC R95021:
DE 02-JUL-1996 (first entry)
DE Hepatitis GB virus (HGBV) contig C protein prod.
KW Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine;
KW reagents; non-A; non-B; non-C; non-D; non-E; clone; GB contig C;
KW tamarin; infected plasma; lambda phage; cDNA library.
OS Hepatitis GB virus.
UN Key Location/Qualifiers
FT misc_difference 1..3011
FT /note="Others correspond to degenerate or STOP
FT codons in T04247"
EN WO9521922.A2
PD 17-AUG-1995.
PE 14-FEB-1995; U02118.
PR 14-FEB-1994; US-196030.
PR 13-MAY-1994; US-242654.
PR 29-JUL-1994; US-283314.
PR 23-NOV-1994; US-344190.
PR 23-NOV-1994; US-344185.
PR 27-JAN-1995; US-344557.
PA (ABBO-) ABBOTT LAB.
PI Buljk SL, Dawson GJ, Desai SM, Erker JC, Leary TP,
PI Muerhoff AS, Mushahwar IK, Pilot-Matias TJ, Schlauder GG;
PI Simons JN;
DR WPI: 95-293123/38.
N-PSDB: T04247.
PT Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful

```

```

PT for diagnosis and therapy of hepatitis GB virus
PS Example 18: Pages 483-496; 661pp; English.
CC Double stranded hepatitis GB virus (HGBV) DNA obtd. from HGBV
CC infected tamarin plasma, using standard procedures, was used to
CC prepare a lambda phage HGBV cDNA library. Clones were rescued
CC from the lambda phage, searched against a sequence database and
CC found to be unique HGBV sequences. The clones were then used to
CC assemble the sequences GB contig A and B, which were amplified
CC using random primers. The prod. of which was amplified to give a
CC fragment of GB contig C, then using specified primers the
CC complete sequence of GB contig C was assembled to give T04247,
CC which encodes the proteins R95020-21 (the 3 possible coding strand
CC reading frames). Reagents which comprise the HGBV DNA, or its
CC protein prods. can be used for the diagnosis, therapy or in a
CC vaccine to prevent HGBV infection.
SO Sequence 3011 AA;

```

```

Query Match 64.0%; Score 48; DB 16; Length 3011;
Best Local Similarity 55.6%; Pred. No. 2.10e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 1923 fpmfyt1 1931
OY 1 fpmfyt1 9

```

Search completed: Fri Sep 11 12:58:00 1998  
Job time : 17 secs.



J.F.; McDonald, L.; Uterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 #journal Nature (1997) 390:364-370  
 #title The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.  
 #cross-references MUID:98049343  
 #accession C69293  
 #status preliminary; nucleic acid sequence not shown;  
 #molecule\_type DNA  
 #residues 1-319 #label KLE  
 #cross-references GB:AE000782; TIGR:AF0347  
 #length 319 #molecular-weight 35267 #checksum 7060

Query Match 74.7%; Score 56; DB 2; Length 319;  
 Best Local Similarity 66.7%; Pred. No. 5.44e+00;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 68 FVGNFYPL 76  
 1 1 1 1 1 1  
 1 FAMPNFYTL 9

RESULT 3  
 ENTRY 3  
 TITLE A65050 #type complete  
 membrane-bound lytic transglycosylase (EC 3.2.1.-) B  
 precursor - *Escherichia coli*  
 ORGANISM *Escherichia coli*  
 DATE 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 13-Mar-1998  
 A65050; S65868; S77642  
 A64720

ACCESSIONS  
 REFERENCE  
 #authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
 #journal Science (1997) 277:1453-1462  
 #title The complete genome sequence of *Escherichia coli* K-12.  
 #cross-references MUID:9742617  
 #accession A65050  
 #status nucleic acid sequence not shown; translation not shown  
 #molecule\_type DNA  
 #residues 1-361 #label BLAT  
 #cross-references GB:AE000354; GB:U00096; NID:q2367149; PID:q1789053; UMCP:D2701

REFERENCE  
 #experimental\_source strain K-12, substrain MG1655  
 #accession S65868  
 #authors Dijkstra, A.J.; Hermann, F.; Keck, W.  
 #journal FEBS Lett. (1995) 366:115-118  
 #title Cloning and controlled overexpression of the gene encoding the 35 kDa soluble lytic transglycosylase from *Escherichia coli*.  
 #accession S65868  
 #status preliminary; not compared with conceptual translation  
 #molecule\_type DNA  
 #residues 1-34, 'A', 36-361 #label DJJ  
 #length 37642  
 #authors Ehler, K.; Hoeltje, J.V.; Templin, M.F.  
 #journal Mol. Microbiol. (1995) 16:761-768  
 #title Cloning and expression of a murein hydrolase lipoprotein from *Escherichia coli*.  
 #accession S77642  
 #status preliminary; nucleic acid sequence not shown;  
 #molecule\_type DNA  
 #residues 1-361 #label EHL  
 #cross-references EMBL:U18785  
 #note the nucleotide sequence was submitted to the EMBL Data

Library, December 1994

GENETICS  
 #gene mlb  
 #KEYWORDS glycosidase; hydrolase  
 #FEATURE 1-18  
 #product mlb protein #status experimental #label MAF  
 #length 361 #molecular-weight 40256 #checksum 6127

Query Match 73.3%; Score 55; DB 2; Length 361;  
 Best Local Similarity 55.6%; Pred. No. 8.02e+00;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 327 YGLNPFYTL 335  
 1 1 1 1 1 1  
 1 FAMPNFYTL 9

RESULT 4  
 ENTRY 4  
 TITLE A69927 #type complete  
 ribonucleoside-diphosphate reductase (alph) homolog ysoo -  
 ORGANISM *Bacillus subtilis*  
 DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Dec-1997  
 A69927  
 A69580

ACCESSIONS  
 REFERENCE  
 #authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bester, M.G.; Bessieres, P.; Bolochin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denzot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Enlian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Chim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandl, G.; Giesepf, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueil, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moesti, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Prescan, E.; Pujic, P.; Punelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rooba, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandendol, M.; Vannier, F.; Vassartoli, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Welter, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

REFERENCE  
 #journal Nature (1997) 390:249-256  
 #title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 #accession A69927  
 #status preliminary; nucleic acid sequence not shown;  
 #molecule\_type DNA  
 #residues 1-786 #label KUN  
 #cross-references EMBL:U18785  
 #experimental\_source strain 168  
 #GENETICS  
 #gene ysoo

```
Db      578 FMGVNFEYSL 586
      1 1 1 1 1 : 1
QY      1 FAMPNFEYTL 9
```

RESULT 5

|          |                                      |   |
|----------|--------------------------------------|---|
| ENTRY    | B69020                               | #LysE complete                                    |
| TITLE    | hypothetical protein MTH15           | - Methanobacterium                                |
| ORGANISM | thermoautotrophicum (strain Delta H) |   |
| DATE     | 05-Dec-1997                          | #formal_name Methanobacterium thermoautotrophicum |
|          | 05-Dec-1997                          | #sequence_revision 05-Dec-1997 #text_change       |

|            |        |
|------------|--------|
| ACCESSIONS | B69020 |
| REFERENCE  | A69000 |
| } #authors | Smith, |

#journal  
#title  
#authors  
Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H.,  
Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D.,  
Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, E.  
Lumm, W., Pochiert, B., Qiu, D., Spedtorra, R., Victor, F.  
Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso  
A.; Bush, D.; Safer, H.; Patwell, D.; Prabakar, S.;  
McDonough, S.; Shimer, G.; Goyal, A.; Pietrowski, S.;  
Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,  
J.; Reeve, J.N.  
J. Bacteriol. (1997) 179:7135-7155  
Complete genome sequence of *Meharobacterium*

thermoautotrophicum Delta H: functional analysis and comparative genomics.  
#cross-references MUID:98037514  
#accession B69020

```
##status      preliminary; nucleic acid sequence not shown
              translation not shown
```

```

#molecule_type DNA
##residues 1-189 ##label MTH
##cross-references GB:AE000666
##experimental_source strain Delta H

```

|                   |        |
|-------------------|--------|
| GENETICS          |        |
| #gene             | MTG115 |
| #start_codon      | TTG    |
| SUMMARY           |        |
| #length           | 189    |
| #molecular-weight | 21668  |
| #checksum         | 27322  |

|                        |        |                     |       |                                 |
|------------------------|--------|---------------------|-------|---------------------------------|
| Query Match            | 72.0%; | Score 54;           | DB 2; | length 189;                     |
| 'Best Local Similarity | 66.7%; | Pred. No. 1.18e+01; |       |                                 |
| Matches                | 6;     | Conservative        | 2;    | Mismatches 1; Indels 0; Gaps 0; |

```

Db      100 FTMPSSTL 108
      1:1:1:1:1
QY      1 FAMPNFTL 9

```

## RESULT 6

| ENTRY                          | #type complete                               |
|--------------------------------|--|
| TITLE                          |  |
| ALTERNATE_NAMES                |  |
| ORGANISM                       |  |
| DATE                           |  |
| S63383                         | Yeast ( <i>Saccharomyces cerevisiae</i> )    |
| POP2 protein                   |  |
| protein N3470; protein YNR052c |  |
| formal_name                    | <i>Saccharomyces cerevisiae</i>              |
| 21-Apr-1996                    | #sequence, revision 03-May-1996 #text change |

06-Feb-1998  
ACCESSIONS S63383, S35997, S35996, S36939, S27438  
REFERENCE S63346  
#authors Pohl, T.M.

```
#submission submitted to the Protein Sequence Database, April 1996
#accession 563383
#molecule_type DNA
#residues 1-433 ##label POH
```

```
##cross_references EMBL:Z71667; NID:q1302567; PID:e2396839; PID:q1302568;
MIPS:YNN052C
##experimental_source strain S286C
REFERENCE S35996
```

```
##status      nucleic acid sequence not shown
##molecule-type  DNA
##residues      1-80,82-411,'M',413-433 ##label
##cross-references  GB:DI2807
##experimental source strain  S288C
#accession      S35596
```

```
##status      nucleic acid sequence not shown
##molecule_type  DNA
##residues      1-40, 'O', 42-91, '0000000000000000', 93-111, 117-277, 'S',
                279-433 ##label SAM
```

```
##cross-references GB:DL2808
##experimental_source strain A364A
REFERENCE S36929
```

```

#authors      Sakai, A.; Chiriacanur, T.; Shimizu, Y.; Hishinuma, F.
#submission   submitted to the EMBL Data Library, August 1992
#accession    S36929
##molecule_type DNA
##residues    1-91, '0000000000000000', 92-111, 117-277, 'S', 279-433

```

```
##label SA2
##cross-references GB:D12808; NID:g218462; PID:d1002742; PID:g218463
##experimental_source strain A364A
REFERENCE
S27437
```

|             |  |
|-------------|--|
| #authors    | Cusick, M.E.                                   |
| #submission | submitted to the EMBL Data Library, March 1992 |
| #accession  | S27438   |
| #molecule   | type DNA                                       |

```

##residues      213-433  ##label CUS
##cross-references EMBL:M88607; NID:g172079; PID:g172080
NETICS

```

```
#gene          SGD:POP2; CAF1
#cross-references SGD:S0005335; MIPS:YNR052c
#map_position 14R
#array         #length 433 #molecular-weight 49682 #checksum 261
```

|            |       |                     |       |               |
|------------|-------|---------------------|-------|---------------|
| Very Match | 72.0% | Score 54;           | DB 2; | Length 433;   |
| Best Local | 85.7% | Pred. No. 1.18e+01; |       |               |
| Matches    | 6;    | Conservative        | 1;    | Mismatches 0; |
|            |       |                     |       | Gaps 0;       |

```

338 MPNEYDL 344
      |||||
3  MPNEYTL 9

```

RESULT 7 S11

FILE  
#####  
#life\_classification  
neurofibromatosis-related protein NF1 - mouse (fragment  
#normal\_name Mus musculus #common\_name house mouse  
18-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change  
10-Nov-1995

|           |  |
|-----------|--|
| SESSIONS  | 10 NOV 1955  |
| REFERENCE | S11510   |
| #authors  | S11510   |
|           | Butcher, A.M.; Cleveland, L.S.; Jenkins, N.A.; Copeland<br>N |

| #journal<br>#title  | N.G. |
|---|------|
| Nature (1990) 347:291-294   |      |
| Sequence homology shared by neurofibromatosis type-1 gene and<br>IRA-1 and IRA-2 negative regulators of the RAS cyclic AMP<br>pathway |      |

```

#cross-references MUID:903845695
#accession S11510
#status preliminary
#relat...

```

```

# #molecule_type mRNA
# #residues 1-621 #label BUC
# #length 621 #checksum 58
SUMMARY

```

```

Query Match      72.0%; Score 54; DB 2; Length 621;
Best Local Similarity 55.6%; Pred. No. 1.18e+01;
Matches          5; Conservative 3; Mismatches 1; Indels 0; Gaps 0

```

DB 209 FSLPKFTL 217  
 OY 1 FAMPNFTL 9

RESULT 8  
 ENTRY 572001 #type complete  
 TITLE dolichyl-phosphate-mannose--glycolipid  
 (alpha-mannosyltransferase (EC 2.4.1.130) PMT3 - yeast  
 (Saccharomyces cerevisiae)  
 ORGANISM #formal name Saccharomyces cerevisiae  
 #accession 12-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change  
 DATE 13-Mar-1998

ACCESSIONS  
 #authors Pearson, B.M.; Hernandez, Y.; Payne, J.; Wolf, S.S.;  
 #journal Yeast (1996) 12:1021-1031  
 #title Sequencing of a 35.71 kb DNA segment on the right arm of  
 yeast chromosome XV reveals regions of similarity to  
 chromosomes I and XIII.

#accession S72001  
 #status nucleic acid sequence not shown; translation not shown  
 #molecule\_type DNA  
 #residues 1-753 #label PEM  
 #cross-references EMBL:X80365  
 #note the nucleotide sequence was submitted to the EMBL Data  
 Library, August 1995

GENETICS  
 #gene PMT3  
 #keywords glycosyltransferase; hexosyltransferase  
 SUMMARY #length 753 #molecular\_weight 86322 #checksum 3415

Query Match 72.0%; Score 54; DB 2; Length 753;  
 Best Local Similarity 55.6%; Pred. No. 1.18e+01;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 641 FVMAFYPL 649  
 OY 1 FAMPNFTL 9

RESULT 9  
 ENTRY S58331 #type complete  
 TITLE dolichyl-phosphate-mannose--protein mannosyltransferase (EC  
 2.4.1.109) PMT3 - yeast (Saccharomyces cerevisiae)  
 ALTERNATE\_NAMES protein O6148; protein YOR321w  
 ORGANISM #formal name Saccharomyces cerevisiae  
 DATE 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change  
 06-Feb-1998

ACCESSIONS  
 #authors Pearson, B.M.; Hernandez, Y.; Wolf, S.S.; Kalogeropoulos, A.;  
 #journal Yeast (1995) 11:1345-1351  
 #title PMT3 and PMT4, two new members of the  
 protein-O-mannosyltransferase gene family of Saccharomyces  
 cerevisiae.

#accession S58331  
 #status nucleic acid sequence not shown  
 #molecule\_type DNA  
 #residues 1-753 #label PEM  
 #cross-references EMBL:X90565; NID:g940836; PID:g940852  
 #accession S60414

REFERENCE  
 #authors Immervoll, T.; Gentzsch, M.; Tanner, W.  
 #journal Yeast (1995) 11:1345-1351  
 #title PMT3 and PMT4, two new members of the  
 protein-O-mannosyltransferase gene family of Saccharomyces  
 cerevisiae.

#accession S60414  
 #status nucleic acid sequence not shown  
 #molecule\_type DNA  
 #residues 1-396; 'H', 398-566; 'N', 568-753 #label IMM  
 #cross-references EMBL:X83797; NID:g633651; PID:g633652  
 #accession S67213

#authors Pearson, B.M.; Hernandez, Y.; Kalogeropoulos, A.; Schweizer,  
 M.  
 #submission submitted to the Protein Sequence Database, July 1996  
 #accession S67227  
 #molecule\_type DNA  
 #residues 1-753 #label PEM  
 #cross-references EMBL:Z75229; NID:g1420703; PID:e252150; PID:g1420704;  
 #experimental\_source strain S288C

GENETICS  
 #gene SGD:PMT3  
 #cross-references SGD:S0005848; MIPS:YOR321w  
 #map\_position 15R  
 #keywords glycosyltransferase; hexosyltransferase; transmembrane  
 protein

FEATURE  
 55-71  
 166-182 #domain transmembrane #status predicted #label TM1  
 192-208 #domain transmembrane #status predicted #label TM2  
 239-255 #domain transmembrane #status predicted #label TM3  
 284-300 #domain transmembrane #status predicted #label TM4  
 607-623 #domain transmembrane #status predicted #label TM5  
 640-656 #domain transmembrane #status predicted #label TM6  
 704-720 #domain transmembrane #status predicted #label TM7  
 #domain transmembrane #status predicted #label TM8

SUMMARY #length 753 #molecular\_weight 86322 #checksum 3415

Query Match 72.0%; Score 54; DB 2; Length 753;  
 Best Local Similarity 55.6%; Pred. No. 1.18e+01;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 641 FVMAFYPL 649  
 OY 1 FAMPNFTL 9

RESULT 10  
 ENTRY B55282 #type complete  
 TITLE neurofibromatosis-related protein NF1 - human  
 ALTERNATE\_NAMES GTPase activating protein homolog NF1; neurofibromin  
 ORGANISM #formal name Homo sapiens #common\_name man  
 DATE 10-Feb-1995 #sequence\_revision 10-Feb-1995 #text\_change  
 20-Mar-1998

ACCESSIONS  
 #authors Marchuk, D.A.; Saulino, A.M.; Tavakoli, R.; Svaroop, M.;  
 Wallace, M.R.; Andersen, L.B.; Mitchell, A.L.; Gutmann,  
 D.H.; Boguski, M.; Collins, F.S.  
 #journal Genomics (1991) 11:931-940  
 #title cDNA cloning of the type 1 neurofibromatosis gene: complete  
 sequence of the NF1 gene product.

#cross-references MUID:92147138  
 #accession B55282  
 #status not compared with conceptual translation  
 #molecule\_type mRNA  
 #residues 1-2818 #label MAR  
 #cross-references GB:M82814; NID:g189164; PID:g189165  
 #note Sequence extracted from NCBI backbone (NCBIP:80176)

#accession A55282  
 #status preliminary  
 #molecule\_type mRNA  
 #residues 1-334 #label MA2  
 #note Sequence extracted from NCBI backbone (NCBIP:80172)

REFERENCE  
 #authors Xu, G.; O'Connell, P.; Viskochil, D.; Cawthon, R.; Robertson,  
 M.; Culver, M.; Dunn, D.; Stevens, J.; Gesteland, R.;  
 White, R.; Weiss, R.  
 #journal Cell (1990) 62:599-608  
 #title The neurofibromatosis type 1 gene encodes a protein related  
 to GAP.

#cross-references MUID:90335969  
 #accession A35879



```

##status Preliminary
##molecule_type mRNA
##residues 335-495, 'I', 497-1555, 'H', 1556-2818 ##label XUA
##cross-references GB:M38106; GB:M57449; NID:9189169; PID:9189170
REFERENCE
#authors Cavthon, R.M.; Weiss, R.; Xu, G.; Viskochil, D.; Culver, M.;
Stevens, J.; Robertson, M.; Dunn, D.; Gesteland, R.;
#journal O'Connell, P.; White, R.
Cell (1990) 62:193-201
#title A major segment of the neurofibromatosis type 1 gene: cDNA
sequence, genomic structure, and point mutations.
#cross-references MUID:90304909
#accession A35605
##status Preliminary
##molecule_type mRNA
##residues 1585-2687 ##label CAW
##cross-references EMBL:M38107; EMBL:M57449
REFERENCE
#authors Cavthon, R.M.; Weiss, R.; Xu, G.; Viskochil, D.; Culver, M.;
Stevens, J.; Robertson, M.; Dunn, D.; Gesteland, R.;
O'Connell, P.; White, R.
Cell (1990) 62:608D
#accession A35910
##status Preliminary; nucleic acid sequence not shown; not
compared with conceptual translation
##molecule_type mRNA
##residues 2688-2818 ##label CA2
REFERENCE
#authors Wallace, M.R.; Marchuk, D.A.; Andersen, L.B.; Letcher, R.;
Odeh, H.M.; Saulino, A.M.; Fountain, J.W.; Breton, A.;
Nicholson, J.; Mitchell, A.L.; Brownstein, B.H.; Collins,
F.S.
Science (1990) 249:181-186
#title Type 1 neurofibromatosis gene: identification of a large
transcript disrupted in three NF1 patients.
#cross-references MUID:90319792
#accession A35222
##status Preliminary
##molecule_type mRNA
##residues 2209-2818 ##label WML
##cross-references GB:M60496; NID:9189157; PID:9189158; GB:M49193
REFERENCE
#authors Martin, G.A.; Viskochil, D.; Bollag, G.; McCabe, P.C.;
Crossier, W.J.; Haubruck, H.; Conroy, L.; Clark, R.;
O'Connell, P.; Cavthon, R.M.; Innis, W.A.; McCormick, F.
Cell (1990) 63:843-849
#title The GAP-related domain of the neurofibromatosis type 1 gene
product interacts with ras p21.
#cross-references MUID:91029515
#accession A36297
##molecule_type mRNA
##residues 1096-1569, 'TPPEPPT' ##label MA3
##cross-references GB:M61213; NID:9189162; PID:9189163
#note this clone includes an epitope tag at the 3' end
encoding the sequence TPPEPPT, not part of dystrophin
but recognized by the monoclonal antibody KT3
REFERENCE
#authors Nishi, T.; Lee, P.S.; Oka, K.; Levin, V.A.; Tanase, S.;
Morino, Y.; Saye, H.
Oncogene (1991) 6:1555-1559
#title Differential expression of two types of the neurofibromatosis
type 1 (NF1) gene transcripts related to neuronal
differentiation.
#cross-references MUID:92019823
#accession I58356
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1168-1545 ##label RES
##cross-references GB:M60915; NID:9189159; PID:9189160
GENETICS
#gene GDB:NFI
##cross-references GDB:120231; OMIM:162200
#map_position 17q11.2-17q11.2

```

```

#introns 1370/3
#note the list of introns is incomplete
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
KEYWORDS #alternative splicing; tumor suppressor
FEATURE
1235-1449
SUMMARY #domain ras-specific GAP catalytic domain homology
#label GAP
#length 2818 #molecular-weight 317030 #checksum 2858
Query Match 72.0%; Score 54; DB 2; Length 2818;
Best Local Similarity 55.6%; Pred. No. 1.18e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 2135 FSLPKFYLL 2143
1 FAMPNFYLL 9
RESULT 11
ENTRY JC5196 #type complete
TITLE neurofibromin I - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change
10-Sep-1997
ACCESSIONS JC5196
REFERENCE JC5196
#authors Suzuki, H.; Takahashi, K.; Yasumoto, K.; Fuse, N.; Shibahara,
S.
J. Biochem. (1996) 120:1048-1054
#title Differential tissue-specific expression of neurofibromin
isoform mRNAs in rat.
#accession JC5196
##status Preliminary; nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-2820 ##label SUZ
#cross-references DDBJ:D45201; NID:91841313; PID:91841314
COMMENT This protein contains a GTPase-activating protein-related domain
which is responsible for the stimulatory effect of neurofibromin
on the tyrosinase promoter activity.
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
FEATURE 1177-1436
#domain GTPase-activating protein related #status
1237-1451 #predicted #label GRD
#domain ras-specific GAP catalytic domain homology
#label GAP
SUMMARY #length 2820 #molecular-weight 317080 #checksum 6628
Query Match 72.0%; Score 54; DB 2; Length 2820;
Best Local Similarity 55.6%; Pred. No. 1.18e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 2137 FSLPKFYLL 2145
1 FAMPNFYLL 9
RESULT 12
ENTRY I54352 #type fragment
TITLE neurofibromin - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
25-Apr-1997
ACCESSIONS I54352
REFERENCE I54352
#authors Bernards, A.; Smolthers, A.J.; Hannigan, G.E.; Murthy, A.E.;
Gusella, J.F.
Hum. Mol. Genet. (1993) 2:645-650
#title Mouse neurofibromatosis type 1 cDNA sequence reveals high
degree of conservation of both coding and non-coding mRNA
segments.
#cross-references MUID:93357730
#accession I54352
##status Preliminary; translated from GB/EMBL/DBJ

```



\*\*\*\*\*  
 M P S E R E H  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Sep 11 12:59:06 1998; Maspar time 2.36 Seconds  
 Tabular output not generated.

Title: >US-08-452-843-6  
 Description: (1-9) from US08452843.pep  
 Perfect score: 75  
 Sequence: 1 FAMMPFVTL 9

Scoring table:  
 PAM 150  
 Gap 15

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot35  
 1:swiss1

Statistics: Mean 25.095; Variance 33.558; scale 0.748

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description                        | Pred. No. |
|------------|-------|-------------|--------|----|------------------------------------|-----------|
| 1          | 56    | 74.7        | 195    | 1  | YEH7_YEAST HYPOTHETICAL 21.7 KD P  | 1.54e+00  |
| 2          | 55    | 73.3        | 361    | 1  | MLTB_ECOLI MEMBRANE-BOUND LYTIC M  | 2.38e+00  |
| 3          | 54    | 72.0        | 270    | 1  | NFL1_CHICK NEUROFILBROMIN (NEUROFI | 3.68e+00  |
| 4          | 54    | 72.0        | 433    | 1  | POP2_YEAST POR2 PROTEIN.           | 3.68e+00  |
| 5          | 54    | 72.0        | 753    | 1  | PMT3_YEAST DOLICHYL-PHOSPHATE-MAN  | 3.68e+00  |
| 6          | 54    | 72.0        | 2839   | 1  | NFL1_HUMAN NEUROFILBROMIN (NEUROFI | 3.68e+00  |
| 7          | 54    | 72.0        | 2841   | 1  | NFL1_MOUSE NEUROFILBROMIN (NEUROFI | 3.68e+00  |
| 8          | 52    | 69.3        | 393    | 1  | YNBS_YEAST HYPOTHETICAL 44.5 KD P  | 8.63e+00  |
| 9          | 51    | 68.0        | 354    | 1  | TYRA_LACIA PREPHENATE DEHYDROGENA  | 1.31e+01  |
| 10         | 51    | 68.0        | 869    | 1  | AMPN_ECOLI AMINOPEPTIDASE N (EC 3  | 1.31e+01  |
| 11         | 50    | 66.7        | 169    | 1  | NNGC_MARPO NADH-PLASTOQUINONE OXI  | 1.97e+01  |
| 12         | 50    | 66.7        | 198    | 1  | PEMT_RAT PHOSPHATIDYLETHANOLAMIT   | 1.97e+01  |
| 13         | 50    | 66.7        | 246    | 1  | SPT1_BACSU SPORE COAT POLYSACCHAR  | 1.97e+01  |
| 14         | 50    | 66.7        | 381    | 1  | Y926_HELPY HYPOTHETICAL 47.6 KD P  | 1.97e+01  |
| 15         | 50    | 66.7        | 410    | 1  | Y928_CABEL HYPOTHETICAL 47.6 KD P  | 1.97e+01  |
| 16         | 50    | 66.7        | 997    | 1  | YNM3_YEAST HYPOTHETICAL 110.9 KD   | 1.97e+01  |
| 17         | 49    | 65.3        | 700    | 1  | RIR1_BACSU RIBONUCLEOSIDE-DIPHOSP  | 2.96e+01  |
| 18         | 49    | 65.3        | 728    | 1  | MYBA_XENIA MYB-RELATED PROTEIN A   | 2.96e+01  |
| 19         | 49    | 65.3        | 844    | 1  | HEXA_STRPN DNA MISMATCH REPAIR PR  | 2.96e+01  |
| 20         | 48    | 64.0        | 375    | 1  | ACT_GIALA ACTIN                    | 4.41e+01  |
| 21         | 48    | 64.0        | 402    | 1  | PA11_BOVIN PLASMINOGEN ACTIVATOR   | 4.41e+01  |
| 22         | 48    | 64.0        | 463    | 1  | UHPT_ECOLI HEXOSE PHOSPHATE TRANS  | 4.41e+01  |
| 23         | 48    | 64.0        | 463    | 1  | UHPT_SALTY HEXOSE PHOSPHATE TRANS  | 4.41e+01  |

|    |    |      |      |   |                                    |          |
|----|----|------|------|---|------------------------------------|----------|
| 24 | 48 | 64.0 | 468  | 1 | YOPH_YERPS PROTEIN-TYROSINE PHOSP  | 4.41e+01 |
| 25 | 48 | 64.0 | 468  | 1 | YOPH_YEREN PROTEIN-TYROSINE PHOSP  | 4.41e+01 |
| 26 | 48 | 64.0 | 477  | 1 | PEN3_ADECC PENTON PROTEIN (VIRION  | 4.41e+01 |
| 27 | 48 | 64.0 | 543  | 1 | IEFS_HUMAN TRANSFORMATION-SENSITI  | 4.41e+01 |
| 28 | 48 | 64.0 | 758  | 1 | PMT2_YEAST DOLICHYL-PHOSPHATE-MAN  | 4.41e+01 |
| 29 | 48 | 64.0 | 3005 | 1 | POLG_TVAV GENOME POLYPROTEIN (CO   | 4.41e+01 |
| 30 | 47 | 62.7 | 122  | 1 | YBHC_ECOLI HYPOTHETICAL 14.2 KD P  | 6.52e+01 |
| 31 | 47 | 62.7 | 219  | 1 | CAT_ECOLI CHLORAMPHENICOL ACETYL   | 6.52e+01 |
| 32 | 47 | 62.7 | 246  | 1 | Y181_METUA HYPOTHETICAL PROTEIN M  | 6.52e+01 |
| 33 | 47 | 62.7 | 253  | 1 | PHY3_MASIA PHYCOBILISOME ROD-CORE  | 6.52e+01 |
| 34 | 47 | 62.7 | 355  | 1 | PURS_ARATH PHOSPHORIBOSYLFORMYLGL  | 6.52e+01 |
| 35 | 47 | 62.7 | 377  | 1 | PYRC_ARATH DIHYDROOCTASE PRECURS   | 6.52e+01 |
| 36 | 47 | 62.7 | 402  | 1 | PAIL_RAT PLASMINOGEN ACTIVATOR     | 6.52e+01 |
| 37 | 47 | 62.7 | 402  | 1 | PAIL_MOUSE PLASMINOGEN ACTIVATOR   | 6.52e+01 |
| 38 | 47 | 62.7 | 482  | 1 | YPT1_CABEL ATP-DEPENDENT PERMEASE  | 6.52e+01 |
| 39 | 47 | 62.7 | 685  | 1 | MDL1_CANAL RAP HOMOLOG SERINE/THR  | 6.52e+01 |
| 40 | 47 | 62.7 | 781  | 1 | KRAF_DROME RAP HOMOLOG SERINE/THR  | 6.52e+01 |
| 41 | 47 | 62.7 | 949  | 1 | YMP9_YEAST PUTATIVE 109.8 KD TRAN  | 6.52e+01 |
| 42 | 46 | 61.3 | 306  | 1 | RCGM_CHLAU REACTION CENTER PROTEI  | 9.58e+01 |
| 43 | 46 | 61.3 | 471  | 1 | SG3_MOUSE SECRETORYGRANIN IIT PREC | 9.58e+01 |
| 44 | 46 | 61.3 | 665  | 1 | YJCS_ECOLI HYPOTHETICAL 73.7 KD P  | 9.58e+01 |
| 45 | 46 | 61.3 | 873  | 1 | PCI_HUMAN PLASMA-CELL MEMBRANE G   | 9.58e+01 |

## ALIGNMENTS

| RESULT ID  | YEH7_YEAST   | STANDARD  | PRT  | 195 AA. |
|--|--|-----------|------|---------|
| AC   | P39978;  |           |      |         |
| DT   | 01-FEB-1995 (REL. 31, CREATED)   |           |      |         |
| DT   | 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)                            |           |      |         |
| DT   | 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)                          |           |      |         |
| DE   | HYPOTHETICAL 21.7 KD PROTEIN IN HXT8-CANI INTERGENIC REGION.           |           |      |         |
| GN   | YEL067C.   |           |      |         |
| OS   | SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).                              |           |      |         |
| OC   | EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.                       |           |      |         |
| RN   | [1]  |           |      |         |
| RP   | SEQUENCE FROM N.A.   |           |      |         |
| RC   | STRAIN-6288C / AB972;  |           |      |         |
| RA   | DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,     |           |      |         |
| RA   | AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,   |           |      |         |
| RA   | CHUNG E., DUNCAN M., GUZMAN C., HARTZELL G., HUNNICE-SMITH S.,         |           |      |         |
| RA   | HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,             |           |      |         |
| RA   | MOSEDALE D., NAKAHARA K., NAMATH A., NORGEN R., OFFNER P., OH C.,      |           |      |         |
| RA   | PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,     |           |      |         |
| RA   | TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.,                 |           |      |         |
| RL   | SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.                   |           |      |         |
| DR   | EMBL: U18795; G603251; -   |           |      |         |
| KW   | HYPOTHETICAL PROTEIN.  |           |      |         |
| SQ   | SEQUENCE 195 AA; 21721 MW; 2EA97A20 CRC32;                             |           |      |         |
| Query Match  |  |           |      |         |
| Best Local Similarity 74.7%; Score 56; DB 1; Length 195;   |  |           |      |         |
| Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0; |  |           |      |         |
| Db   | 17 FDMPTFEVL 25  |           |      |         |
| QY   | 1 FAMMPFVTL 9  |           |      |         |
| RESULT 2   |  |           |      |         |
| ID   | MLTB_ECOLI   | STANDARD; | PRT; | 361 AA. |
| AC   | P41052;  |           |      |         |
| DT   | 01-FEB-1995 (REL. 31, CREATED)   |           |      |         |
| DT   | 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)                            |           |      |         |
| DT   | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)                          |           |      |         |
| DE   | MEMBRANE-BOUND LYTIC MOREIN TRANSGLYCOSYLASE B PRECURSOR (EC 3.2.1.-)  |           |      |         |
| DE   | (MUREIN HYDROLASE B) (35 KD SOLUBLE LYTIC TRANSGLYCOSYLASE) (SLT35).   |           |      |         |
| GN   | MLTB.  |           |      |         |
| OS   | ESCHERICHIA COLI.  |           |      |         |
| OC   | PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS; |           |      |         |
| OC   | ENTEROBACTERIACEAE.  |           |      |         |
| RN   | [1]  |           |      |         |

RP SEQUENCE FROM N.A.  
 RC STRAIN-K12:  
 RX MEDLINE: 96065704.  
 RA EHLERT K., HOELTJE J.-V., TEMPLIN M.F.:  
 RL MOL. MICROBIOL. 16:761-768(1995).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95309413.  
 RA DIKSTRA A.J., HERMANN F., KECK W.:  
 RL PEST LEFT. 366:115-118(1995).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA BLATTNER F.R., PLUNKETT G., III, MAYHEW G.F., PERNA N.T., GLASNER F.D.:  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12:  
 RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONDO A., HORIUCHI T.,  
 RA IKEMOTO K., INADA T., ISONO K., ITOH T., KAWAI K., KASAI H.,  
 RA KASHIMOTO K., KIM S., KITMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,  
 RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,  
 RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,  
 RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.:  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 [5]  
 RP PRELIMINARY SEQUENCE OF 1-91 FROM N.A.  
 RX MEDLINE: 87194727.  
 RA YAMADA M., SAIER M.H. JR.:  
 RL J. BIOL. CHEM. 262:5455-5463(1987).  
 CC -1- FUNCTION: MUREIN-DEGRADING ENZYME. MAY PLAY A ROLE IN RECYCLING  
 CC OF MURPEPTIDES DURING CELL ELONGATION AND/OR CELL DIVISION.  
 CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF THE BETA-1,4-GLYCOSIDIC BOND  
 CC BETWEEN N-ACETYLGLUCOSAMINE ACID AND N-ACETYLGLUCOSAMINE RESIDUES,  
 CC THEREBY CONSERVING THE ENERGY IN A NEWLY SYNTHESIZED  
 CC 1,6-AMPHIDROBOND IN THE MURAMIC ACID RESIDUE.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID  
 CC ANCHOR AND EXPOSED TO THE PERIPLASMIC SIDE (PROBABLE).  
 DR EMBL: U18785; G642538; -;  
 DR EMBL: AE000354; G178053; -;  
 DR EMBL: D90892; G1800087; -;  
 DR EMBL: J02708; -; NOT ANNOTATED\_CDS.  
 DR ECOGENE: EG12699; MLTB.  
 DR PROSITE: PS00013; PROKAR.LIPOPROTEIN: 1.  
 DR CELL WALL, HYDROLASE, GLYCOSIDASE, SIGNAL, LIPOPROTEIN;  
 KW OUTER MEMBRANE; MULTIGENE FAMILY.  
 FT SIGNAL 1 18 PROBABLE.  
 FT CHAIN 19 361 MEMBRANE-BOUND LYTIC MUREIN  
 FT LIPID 19 19 N-ACYL DIGLYCERIDE B.  
 FT SEQUENCE 361 AA: 40256 MW; 93AF5C59 CRC32;  
 SO  
 Query Match 73.3%; Score 55; DB 1; Length 361;  
 Best Local Similarity 55.6%; Pred. No. 2.38e+00;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Db 327 YCLPNEYTL 335  
 Oy 1 FAMPNEYTL 9  
 RESULT 3  
 ID NFL-CHICK STANDARD; PRT: 270 AA.  
 AC P35608;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1) (FRAGMENT).  
 GN NFL.  
 OS GALLUS GALLUS (CHICKEN).  
 OC EURARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;  
 OC GALLIFORMES.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN:  
 RX MEDLINE: 93282908.  
 RA SCHAFER G.L., CLEMENT G., STOCKER K.M., BAIZER L.:  
 RL MOL. CHEM. NEUROPATHOL. 18:267-278(1993).  
 CC -1- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NFL SHOWS GREATER  
 CC AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE  
 CC A REGULATOR OF RAS ACTIVITY.  
 CC -1- SIMILARITY: TO OTHER RAS GTPASE-ACTIVATING PROTEINS.  
 DR EMBL: S62087; G385582; -;  
 DR PROSITE: PS00509; RAS\_GTPASE\_ACTIV\_1; PARTIAL.  
 DR PROSITE: PS50018; RAS\_GTPASE\_ACTIV\_2; PARTIAL.  
 KW GTPASE ACTIVATION.  
 FT NON\_TER 1 1  
 FT SEQUENCE 270 AA: 30753 MW; 653E2C8C CRC32;  
 SQ  
 Query Match 72.0%; Score 54; DB 1; Length 270;  
 Best Local Similarity 55.6%; Pred. No. 3.68e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Db 31 FSLPNEYTL 39  
 Oy 1 FAMPNEYTL 9  
 RESULT 4  
 ID POP2 YEAST STANDARD; PRT: 433 AA.  
 AC P39008;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE POP2 PROTEIN.  
 GN POP2 OR CAP1 OR YNR052C OR N3470.  
 OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EURARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C; AND A364A;  
 RX MEDLINE: 93117094.  
 RA SAKAI A., CHIBAZAKURA T., SHIMIZU Y., HISHIMURA F.:  
 RL NUCLEIC ACIDS RES. 20:6227-6233(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA POHL T.M.:  
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE OF 213-433 FROM N.A.  
 RL SUBMITTED M.E.:  
 CC -1- FUNCTION: REQUIRED FOR GLUCOSE-DEREPRESSION OF GENE EXPRESSION.  
 CC MAY AFFECT TRANSCRIPTION BY INTERACTING WITH OTHER PROTEINS.  
 CC EMBL: D12807; G218465; -;  
 CC EMBL: D12808; G218463; -;  
 CC EMBL: D12809; G218463; -;  
 CC EMBL: D12810; G218463; -;  
 CC EMBL: D12811; G218463; -;  
 CC EMBL: D12812; G218463; -;  
 CC EMBL: D12813; G218463; -;  
 CC EMBL: D12814; G218463; -;  
 CC EMBL: D12815; G218463; -;  
 CC EMBL: D12816; G218463; -;  
 CC EMBL: D12817; G218463; -;  
 CC EMBL: D12818; G218463; -;  
 CC EMBL: D12819; G218463; -;  
 CC EMBL: D12820; G218463; -;  
 CC EMBL: D12821; G218463; -;  
 CC EMBL: D12822; G218463; -;  
 CC EMBL: D12823; G218463; -;  
 CC EMBL: D12824; G218463; -;  
 CC EMBL: D12825; G218463; -;  
 CC EMBL: D12826; G218463; -;  
 CC EMBL: D12827; G218463; -;  
 CC EMBL: D12828; G218463; -;  
 CC EMBL: D12829; G218463; -;  
 CC EMBL: D12830; G218463; -;  
 CC EMBL: D12831; G218463; -;  
 CC EMBL: D12832; G218463; -;  
 CC EMBL: D12833; G218463; -;  
 CC EMBL: D12834; G218463; -;  
 CC EMBL: D12835; G218463; -;  
 CC EMBL: D12836; G218463; -;  
 CC EMBL: D12837; G218463; -;  
 CC EMBL: D12838; G218463; -;  
 CC EMBL: D12839; G218463; -;  
 CC EMBL: D12840; G218463; -;  
 CC EMBL: D12841; G218463; -;  
 CC EMBL: D12842; G218463; -;  
 CC EMBL: D12843; G218463; -;  
 CC EMBL: D12844; G218463; -;  
 CC EMBL: D12845; G218463; -;  
 CC EMBL: D12846; G218463; -;  
 CC EMBL: D12847; G218463; -;  
 CC EMBL: D12848; G218463; -;  
 CC EMBL: D12849; G218463; -;  
 CC EMBL: D12850; G218463; -;  
 CC EMBL: D12851; G218463; -;  
 CC EMBL: D12852; G218463; -;  
 CC EMBL: D12853; G218463; -;  
 CC EMBL: D12854; G218463; -;  
 CC EMBL: D12855; G218463; -;  
 CC EMBL: D12856; G218463; -;  
 CC EMBL: D12857; G218463; -;  
 CC EMBL: D12858; G218463; -;  
 CC EMBL: D12859; G218463; -;  
 CC EMBL: D12860; G218463; -;  
 CC EMBL: D12861; G218463; -;  
 CC EMBL: D12862; G218463; -;  
 CC EMBL: D12863; G218463; -;  
 CC EMBL: D12864; G218463; -;  
 CC EMBL: D12865; G218463; -;  
 CC EMBL: D12866; G218463; -;  
 CC EMBL: D12867; G218463; -;  
 CC EMBL: D12868; G218463; -;  
 CC EMBL: D12869; G218463; -;  
 CC EMBL: D12870; G218463; -;  
 CC EMBL: D12871; G218463; -;  
 CC EMBL: D12872; G218463; -;  
 CC EMBL: D12873; G218463; -;  
 CC EMBL: D12874; G218463; -;  
 CC EMBL: D12875; G218463; -;  
 CC EMBL: D12876; G218463; -;  
 CC EMBL: D12877; G218463; -;  
 CC EMBL: D12878; G218463; -;  
 CC EMBL: D12879; G218463; -;  
 CC EMBL: D12880; G218463; -;  
 CC EMBL: D12881; G218463; -;  
 CC EMBL: D12882; G218463; -;  
 CC EMBL: D12883; G218463; -;  
 CC EMBL: D12884; G218463; -;  
 CC EMBL: D12885; G218463; -;  
 CC EMBL: D12886; G218463; -;  
 CC EMBL: D12887; G218463; -;  
 CC EMBL: D12888; G218463; -;  
 CC EMBL: D12889; G218463; -;  
 CC EMBL: D12890; G218463; -;  
 CC EMBL: D12891; G218463; -;  
 CC EMBL: D12892; G218463; -;  
 CC EMBL: D12893; G218463; -;  
 CC EMBL: D12894; G218463; -;  
 CC EMBL: D12895; G218463; -;  
 CC EMBL: D12896; G218463; -;  
 CC EMBL: D12897; G218463; -;  
 CC EMBL: D12898; G218463; -;  
 CC EMBL: D12899; G218463; -;  
 CC EMBL: D12900; G218463; -;  
 CC EMBL: D12901; G218463; -;  
 CC EMBL: D12902; G218463; -;  
 CC EMBL: D12903; G218463; -;  
 CC EMBL: D12904; G218463; -;  
 CC EMBL: D12905; G218463; -;  
 CC EMBL: D12906; G218463; -;  
 CC EMBL: D12907; G218463; -;  
 CC EMBL: D12908; G218463; -;  
 CC EMBL: D12909; G218463; -;  
 CC EMBL: D12910; G218463; -;  
 CC EMBL: D12911; G218463; -;  
 CC EMBL: D12912; G218463; -;  
 CC EMBL: D12913; G218463; -;  
 CC EMBL: D12914; G218463; -;  
 CC EMBL: D12915; G218463; -;  
 CC EMBL: D12916; G218463; -;  
 CC EMBL: D12917; G218463; -;  
 CC EMBL: D12918; G218463; -;  
 CC EMBL: D12919; G218463; -;  
 CC EMBL: D12920; G218463; -;  
 CC EMBL: D12921; G218463; -;  
 CC EMBL: D12922; G218463; -;  
 CC EMBL: D12923; G218463; -;  
 CC EMBL: D12924; G218463; -;  
 CC EMBL: D12925; G218463; -;  
 CC EMBL: D12926; G218463; -;  
 CC EMBL: D12927; G218463; -;  
 CC EMBL: D12928; G218463; -;  
 CC EMBL: D12929; G218463; -;  
 CC EMBL: D12930; G218463; -;  
 CC EMBL: D12931; G218463; -;  
 CC EMBL: D12932; G218463; -;  
 CC EMBL: D12933; G218463; -;  
 CC EMBL: D12934; G218463; -;  
 CC EMBL: D12935; G218463; -;  
 CC EMBL: D12936; G218463; -;  
 CC EMBL: D12937; G218463; -;  
 CC EMBL: D12938; G218463; -;  
 CC EMBL: D12939; G218463; -;  
 CC EMBL: D12940; G218463; -;  
 CC EMBL: D12941; G218463; -;  
 CC EMBL: D12942; G218463; -;  
 CC EMBL: D12943; G218463; -;  
 CC EMBL: D12944; G218463; -;  
 CC EMBL: D12945; G218463; -;  
 CC EMBL: D12946; G218463; -;  
 CC EMBL: D12947; G218463; -;  
 CC EMBL: D12948; G218463; -;  
 CC EMBL: D12949; G218463; -;  
 CC EMBL: D12950; G218463; -;  
 CC EMBL: D12951; G218463; -;  
 CC EMBL: D12952; G218463; -;  
 CC EMBL: D12953; G218463; -;  
 CC EMBL: D12954; G218463; -;  
 CC EMBL: D12955; G218463; -;  
 CC EMBL: D12956; G218463; -;  
 CC EMBL: D12957; G218463; -;  
 CC EMBL: D12958; G218463; -;  
 CC EMBL: D12959; G218463; -;  
 CC EMBL: D12960; G218463; -;  
 CC EMBL: D12961; G218463; -;  
 CC EMBL: D12962; G218463; -;  
 CC EMBL: D12963; G218463; -;  
 CC EMBL: D12964; G218463; -;  
 CC EMBL: D12965; G218463; -;  
 CC EMBL: D12966; G218463; -;  
 CC EMBL: D12967; G218463; -;  
 CC EMBL: D12968; G218463; -;  
 CC EMBL: D12969; G218463; -;  
 CC EMBL: D12970; G218463; -;  
 CC EMBL: D12971; G218463; -;  
 CC EMBL: D12972; G218463; -;  
 CC EMBL: D12973; G218463; -;  
 CC EMBL: D12974; G218463; -;  
 CC EMBL: D12975; G218463; -;  
 CC EMBL: D12976; G218463; -;  
 CC EMBL: D12977; G218463; -;  
 CC EMBL: D12978; G218463; -;  
 CC EMBL: D12979; G218463; -;  
 CC EMBL: D12980; G218463; -;  
 CC EMBL: D12981; G218463; -;  
 CC EMBL: D12982; G218463; -;  
 CC EMBL: D12983; G218463; -;  
 CC EMBL: D12984; G218463; -;  
 CC EMBL: D12985; G218463; -;  
 CC EMBL: D12986; G218463; -;  
 CC EMBL: D12987; G218463; -;  
 CC EMBL: D12988; G218463; -;  
 CC EMBL: D12989; G218463; -;  
 CC EMBL: D12990; G218463; -;  
 CC EMBL: D12991; G218463; -;  
 CC EMBL: D12992; G218463; -;  
 CC EMBL: D12993; G218463; -;  
 CC EMBL: D12994; G218463; -;  
 CC EMBL: D12995; G218463; -;  
 CC EMBL: D12996; G218463; -;  
 CC EMBL: D12997; G218463; -;  
 CC EMBL: D12998; G218463; -;  
 CC EMBL: D12999; G218463; -;  
 CC EMBL: D13000; G218463; -;  
 CC EMBL: D13001; G218463; -;  
 CC EMBL: D13002; G218463; -;  
 CC EMBL: D13003; G218463; -;  
 CC EMBL: D13004; G218463; -;  
 CC EMBL: D13005; G218463; -;  
 CC EMBL: D13006; G218463; -;  
 CC EMBL: D13007; G218463; -;  
 CC EMBL: D13008; G218463; -;  
 CC EMBL: D13009; G218463; -;  
 CC EMBL: D13010; G218463; -;  
 CC EMBL: D13011; G218463; -;  
 CC EMBL: D13012; G218463; -;  
 CC EMBL: D13013; G218463; -;  
 CC EMBL: D13014; G218463; -;  
 CC EMBL: D13015; G218463; -;  
 CC EMBL: D13016; G218463; -;  
 CC EMBL: D13017; G218463; -;  
 CC EMBL: D13018; G218463; -;  
 CC EMBL: D13019; G218463; -;  
 CC EMBL: D13020; G218463; -;  
 CC EMBL: D13021; G218463; -;  
 CC EMBL: D13022; G218463; -;  
 CC EMBL: D13023; G218463; -;  
 CC EMBL: D13024; G218463; -;  
 CC EMBL: D13025; G218463; -;  
 CC EMBL: D13026; G218463; -;  
 CC EMBL: D13027; G218463; -;  
 CC EMBL: D13028; G218463; -;  
 CC EMBL: D13029; G218463; -;  
 CC EMBL: D13030; G218463; -;  
 CC EMBL: D13031; G218463; -;  
 CC EMBL: D13032; G218463; -;  
 CC EMBL: D13033; G218463; -;  
 CC EMBL: D13034; G218463; -;  
 CC EMBL: D13035; G218463; -;  
 CC EMBL: D13036; G218463; -;  
 CC EMBL: D13037; G218463; -;  
 CC EMBL: D13038; G218463; -;  
 CC EMBL: D13039; G218463; -;  
 CC EMBL: D13040; G218463; -;  
 CC EMBL: D13041; G218463; -;  
 CC EMBL: D13042; G218463; -;  
 CC EMBL: D13043; G218463; -;  
 CC EMBL: D13044; G218463; -;  
 CC EMBL: D13045; G218463; -;  
 CC EMBL: D13046; G218463; -;  
 CC EMBL: D13047; G218463; -;  
 CC EMBL: D13048; G218463; -;  
 CC EMBL: D13049; G218463; -;  
 CC EMBL: D13050; G218463; -;  
 CC EMBL: D13051; G218463; -;  
 CC EMBL: D13052; G218463; -;  
 CC EMBL: D13053; G218463; -;  
 CC EMBL: D13054; G218463; -;  
 CC EMBL: D13055; G218463; -;  
 CC EMBL: D13056; G218463; -;  
 CC EMBL: D13057; G218463; -;  
 CC EMBL: D13058; G218463; -;  
 CC EMBL: D13059; G218463; -;  
 CC EMBL: D13060; G218463; -;  
 CC EMBL: D13061; G218463; -;  
 CC EMBL: D13062; G218463; -;  
 CC EMBL: D13063; G218463; -;  
 CC EMBL: D13064; G218463; -;  
 CC EMBL: D13065; G218463; -;  
 CC EMBL: D13066; G218463; -;  
 CC EMBL: D13067; G218463; -;  
 CC EMBL: D13068; G218463; -;  
 CC EMBL: D13069; G218463; -;  
 CC EMBL: D13070; G218463; -;  
 CC EMBL: D13071; G218463; -;  
 CC EMBL: D13072; G218463; -;  
 CC EMBL: D13073; G218463; -;  
 CC EMBL: D13074; G218463; -;  
 CC EMBL: D13075; G218463; -;  
 CC EMBL: D13076; G218463; -;  
 CC EMBL: D13077; G218463; -;  
 CC EMBL: D13078; G218463; -;  
 CC EMBL: D13079; G218463; -;  
 CC EMBL: D13080; G218463; -;  
 CC EMBL: D13081; G218463; -;  
 CC EMBL: D13082; G218463; -;  
 CC EMBL: D13083; G218463; -;  
 CC EMBL: D13084; G218463; -;  
 CC EMBL: D13085; G218463; -;  
 CC EMBL: D13086; G218463; -;  
 CC EMBL: D13087; G218463; -;  
 CC EMBL: D13088; G218463; -;  
 CC EMBL: D13089; G218463; -;  
 CC EMBL: D13090; G218463; -;  
 CC EMBL: D13091; G218463; -;  
 CC EMBL: D13092; G218463; -;  
 CC EMBL: D13093; G218463; -;  
 CC EMBL: D13094; G218463; -;  
 CC EMBL: D13095; G218463; -;  
 CC EMBL: D13096; G218463; -;  
 CC EMBL: D13097; G218463; -;  
 CC EMBL: D13098; G218463; -;  
 CC EMBL: D13099; G218463; -;  
 CC EMBL: D13100; G218463; -;  
 CC EMBL: D13101; G218463; -;  
 CC EMBL: D13102; G218463; -;  
 CC EMBL: D13103; G218463; -;  
 CC EMBL: D13104; G218463; -;  
 CC EMBL: D13105; G218463; -;  
 CC EMBL: D13106; G218463; -;  
 CC EMBL: D13107; G218463; -;  
 CC EMBL: D13108; G218463; -;  
 CC EMBL: D13109; G218463; -;  
 CC EMBL: D13110; G218463; -;  
 CC EMBL: D13111; G218463; -;  
 CC EMBL: D13112; G218463; -;  
 CC EMBL: D13113; G218463; -;  
 CC EMBL: D13114; G218463; -;  
 CC EMBL: D13115; G218463; -;  
 CC EMBL: D13116; G218463; -;  
 CC EMBL: D13117; G218463; -;  
 CC EMBL: D13118; G218463; -;  
 CC EMBL: D13119; G218463; -;  
 CC EMBL: D13120; G218463; -;  
 CC EMBL: D13121; G218463; -;  
 CC EMBL: D13122; G218463; -;  
 CC EMBL: D13123; G218463; -;  
 CC EMBL: D13124; G218463; -;  
 CC EMBL: D13125; G218463; -;  
 CC EMBL: D13126; G218463; -;  
 CC EMBL: D13127; G218463; -;  
 CC EMBL: D13128; G218463; -;  
 CC EMBL: D13129; G218463; -;  
 CC EMBL: D13130; G218463; -;  
 CC EMBL: D13131; G218463; -;  
 CC EMBL: D13132; G218463; -;  
 CC EMBL: D13133; G218463; -;  
 CC EMBL: D13134; G218463; -;  
 CC EMBL: D13135; G218463; -;  
 CC EMBL: D13136; G218463; -;  
 CC EMBL: D13137; G218463; -;  
 CC EMBL: D13138; G218463; -;  
 CC EMBL: D13139; G218463; -;  
 CC EMBL: D13140; G218463; -;  
 CC EMBL: D13141; G218463; -;  
 CC EMBL: D13142; G218463; -;  
 CC EMBL: D13143; G218463; -;  
 CC EMBL: D13144; G218463; -;  
 CC EMBL: D13145; G218463; -;  
 CC EMBL: D13146; G218463; -;  
 CC EMBL: D13147; G218463; -;  
 CC EMBL: D13148; G218463; -;  
 CC EMBL: D13149; G218463; -;  
 CC EMBL: D13150; G218463; -;  
 CC EMBL: D13151; G218463; -;  
 CC EMBL: D13152; G218463; -;  
 CC EMBL: D13153; G218463; -;  
 CC EMBL: D13154; G218463; -;  
 CC EMBL: D13155; G218463; -;  
 CC EMBL: D13156; G218463; -;  
 CC EMBL: D13157; G218463; -;  
 CC EMBL: D13158; G218463; -;  
 CC EMBL: D13159; G218463; -;  
 CC EMBL: D13160; G218463; -;  
 CC EMBL: D13161; G218463; -;  
 CC EMBL: D13162; G218463; -;  
 CC EMBL: D13163; G218463; -;  
 CC EMBL: D13164; G218463; -;  
 CC EMBL: D13165; G218463; -;  
 CC EMBL: D13166; G218463; -;  
 CC EMBL: D13167; G218463; -;  
 CC EMBL: D13168; G218463; -;  
 CC EMBL: D13169; G218463; -;  
 CC EMBL: D13170; G218463; -;  
 CC EMBL: D13171; G218463; -;  
 CC EMBL: D13172; G218463; -;  
 CC EMBL: D13173; G218463; -;  
 CC EMBL: D13174; G218463; -;  
 CC EMBL: D13175; G218463; -;  
 CC EMBL: D13176; G218463; -;  
 CC EMBL: D13177; G218463; -;  
 CC EMBL: D13178; G218463; -;  
 CC EMBL: D13179; G218463; -;  
 CC EMBL: D13180; G218463; -;  
 CC EMBL: D13181; G218463; -;  
 CC EMBL: D13182; G218463

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 338 MPNFDL 344  
 OY 3 MPNFDL 9

RESULT 5  
 ID PMT3 YEAST STANDARD; PRT; 753 AA.  
 AC P47190;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE DOLICHTL-PROSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE 3  
 DE (EC: 2.4.1.109)  
 GN PMT3 OR YOR321H OR O6148.  
 OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96158055.  
 RA IMMERVOILL T., GENTZSCH M., TANNER W.;  
 RL YEAST 11:1345-1351(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-S288C / FY1679;  
 RL MEDLINE: 97051589.  
 RA PEARSON B.M., HERNANDO Y., PAYNE J., WOLF S.S., KALOGEROPOULOS A.,  
 RA SCHWEIZER M.;  
 RL YEAST 12:1021-1031(1996).  
 CC -1- FUNCTION: TRANSFERS MANNOSE FROM DOL-P-MANNOSE TO SER OR THR  
 CC RESIDUES ON PROTEINS (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: DOLICHTL PHOSPHATE D-MANNOSE + PROTEIN -  
 CC DOLICHTL PHOSPHATE + O-D-MANNOSYL-PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC  
 CC RETICULUM (BY SIMILARITY).  
 CC -1- SIMILARITY: TO OTHER YEAST PMT.  
 CC EMBL: X83797: G633652: -  
 DR EMBL: X90565: G940852: -  
 DR EMBL: Z75229: E252150: -  
 DR SGD: L0002622: PMT3.  
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; GLYCOPROTEIN; TRANSMEMBRANE;  
 KW ENDOPLASMIC RETICULUM; MULTIGENE FAMILY.  
 FT TRANSMEM 51 71  
 FT TRANSMEM 149 169  
 FT TRANSMEM 175 195  
 FT TRANSMEM 236 256  
 FT TRANSMEM 283 303  
 FT TRANSMEM 603 623  
 FT TRANSMEM 640 660  
 FT TRANSMEM 666 686  
 FT TRANSMEM 704 724  
 FT CARBOHYD 48 48  
 FT CARBOHYD 124 124  
 FT CARBOHYD 324 324  
 FT CARBOHYD 398 398  
 FT CONFLICT 397 397  
 FT CONFLICT 567 567  
 FT SEQUENCE 753 AA; 86329 MM; 37CE9258 CRC32;

Query Match 72.0%; Score 54; DB 1; Length 753;  
 Best Local Similarity 55.6%; Pred. No. 3.68e+00;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 641 FVMAFYPL 649  
 OY 1 FVMAFYPL 9

RESULT 6  
 ID NF1\_HUMAN STANDARD; PRT; 2839 AA.  
 AC P21359;  
 DT 01-MAY-1991 (REL. 18, CREATED)

DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).  
 GN NF1.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE OF 1-1370 AND 1392-2839 FROM N.A.  
 RX MEDLINE: 92147138.  
 RA MARCHUK D.A., SAULINO A., TAVAKOL R., SWAROOP M., WALLACE M.R.,  
 RA ANDERSEN L.B., MITCHELL A.L., GUTMANN D.H., BOGUSKI M., COLLINS F.S.;  
 RL GENOMICS 11:991-940(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 93090270.  
 RA BERNARDS A., HAASE V.H., MURTHY A.E., MENON A., HANNIGAN G.E.,  
 RA GUSSELLA J.F.;  
 RL DNA CELL BIOL. 11:727-734(1992).  
 RN [3]  
 RP SEQUENCE OF 335-1370 AND 1392-2839 FROM N.A.  
 RX MEDLINE: 90335969.  
 RA XU G., O'CONNELL P., VISKOCHIL D., CAWTHON R., ROBERTSON M.,  
 RA CULVER M., DUNN D., STEVENS J., GESTELAND R., WHITE R., WEISS R.;  
 RL CELL 62:599-608(1990).  
 RN [4]  
 RP SEQUENCE OF 1096-1370 AND 1372-1590 FROM N.A.  
 RX MEDLINE: 91029515.  
 RA MARTIN G.A., VISKOCHIL D., BOLLAG G., MCCABE P.C., CROSIER W.J.,  
 RA HAUBRUCK H., CONROY L., CLARK R., O'CONNELL P., CAWTHON R.M.,  
 RA INNIS M., MCCORMICK F.;  
 RL CELL 63:843-849(1990).  
 RN [5]  
 RP SEQUENCE OF 1606-2709 FROM N.A., AND VARIANT PRO-1953.  
 RX MEDLINE: 90304909.  
 RA CAWTHON R.M., WEISS R., XU G., VISKOCHIL D., CULVER M., STEVENS J.,  
 RA ROBERTSON M., DUNN D., GESTELAND R., O'CONNELL P., WHITE R.;  
 RL CELL 62:193-201(1990).  
 RN [6]  
 RP SEQUENCE OF 2230-2839 FROM N.A.  
 RX MEDLINE: 90319792.  
 RA WALLACE M.R., MARCHUK D.A., ANDERSEN L.B., LETCHER R., ODEH H.M.,  
 RA SAULINO A.M., FOUNTAIN J.W., BREKTON A., NICHOLSON J., MITCHELL A.L.,  
 RA BROWNSTEIN B.H., COLLINS F.S.;  
 RL SCIENCE 249:181-186(1990).  
 RN [7]  
 RP ERRATUM.  
 RX MEDLINE: 91102559.  
 RA WALLACE M.R., MARCHUK D.A., ANDERSEN L.B., COLLINS F.S.;  
 RL SCIENCE 250:1749-1749(1990).  
 RN [8]  
 RP SEQUENCE OF 1168-1566 FROM N.A.  
 RX MEDLINE: 92019823.  
 RA NISHI T., LEE P.S., OKA K., LEVIN V.A., TANASE S., MORINO Y.,  
 RA SAYA H.;  
 RL ONCOGENE 6:1555-1559(1991).  
 RN [9]  
 RP SEQUENCE OF 1371-1391 FROM N.A.  
 RX MEDLINE: 93109335.  
 RA ANDERSEN L.B., BALLESTER R., MARCHUK D.A., CHANG E., GUTMANN D.H.,  
 RA SAULINO A.M., CANONIS J., WIGLER M., COLLINS F.S.;  
 RL MOL. CELL. BIOL. 13:487-495(1993).  
 RN [10]  
 RP FUNCTION.  
 RX MEDLINE: 91029516.  
 RA BALLESTER R., MARCHUK D., BOGUSKI M.S., SAULINO A., LETCHER R.,  
 RA WIGLER M., COLLINS F.S.;  
 RL CELL 63:851-859(1990).  
 RN [11]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE: 95072625.  
 RA UPADHAYAYA M., SHAW D.J., HARPER P.S.;  
 RL HUM. MUTAT. 4:83-101(1994).

RN [12] REVIEW ON VARIANTS.  
 RP MEDLINE: 96422425.  
 RX HONG SHEN M., HARPER P.S., UPADHYAYA M.;  
 RA J. MED. GENET. 33:2-17(1996).  
 RN [13]  
 RP VARIANT GLU-1444.  
 RX MEDLINE: 92233466.  
 RA LI Y., BOLLAG G., CLARK R., STEVENS J., CONROY L., PUTTS D., WARD K.,  
 RA FRIEDMAN E., SAKOWITZ W., ROBERTSON M., BRADLEY P., MCCORMICK F.,  
 RA WHITE R., CANTHON R.;  
 RL CELL. 69:275-281(1992).  
 RN [14]  
 RP VARIANTS MET-2164 AND ASN-2192.  
 RX MEDLINE: 93258316.  
 RA UPADHYAYA M., SHEN M., CHERRYSON A., FARNHAM J., MAYNARD J.,  
 RA HUSON S.M., HARPER P.S.;  
 RL HUM. MOL. GENET. 1:735-740(1992).  
 RN [15]  
 RP VARIANT HIS-1721-LEU-1733 DUPLICATION.  
 RX MEDLINE: 93304433.  
 RA TASSABEHI M., STRACHAN T., SHARLAND M., COLLEY A., DONNAI D.,  
 RA HARRIS R., THAKKER N.;  
 RL AM. J. HUM. GENET. 53:90-95(1993).  
 RN [16]  
 RP VARIANT MET-991 DEL.  
 RX MEDLINE: 94108439.  
 RA SHEN M.H., HARPER P.S., UPADHYAYA M.;  
 RL HUM. MOL. GENET. 2:1861-1864(1993).  
 RN [17]  
 RP VARIANT NF1 ASN-2387-PHE-2388 DEL.  
 RX MEDLINE: 94362704.  
 RA ABERNATHY C.R., COLMAN S.D., KOUSSEFF B.G., WALLACE M.R.;  
 RL HUM. MUTAT. 3:347-352(1994).  
 RN [18]  
 RP VARIANT NF1 ALA-2631.  
 RX MEDLINE: 96091873.  
 RA UPADHYAYA M., MAYNARD J., OSBORN M., HUSON S.M., PONDER M.,  
 RA PONDER B.A.J., HARPER P.S.;  
 RL J. MED. GENET. 32:706-710(1995).  
 RN [19]  
 RP VARIANT NF1 ARG-629.  
 RX MEDLINE: 96431167.  
 RA GASPARINI P., D'AGRUMA L., DE CILLIS G.P., BALESTRAZZI P.,  
 RA MINGARELLI R., ZELANTE L.;  
 RL HUM. GENET. 97:492-495(1996).  
 RN [20]  
 RP VARIANT LS ARG-1035.  
 RX MEDLINE: 96400960.  
 RA WU R., LEGIUS E., ROBBRECHT W., DUMOULIN M., CASSIMAN J.-J.,  
 RA FRYNS J.-P.;  
 RL HUM. MUTAT. 8:51-56(1996).  
 RN [21]  
 RP VARIANTS NF1 ARG-844 AND PRO-898.  
 RX MEDLINE: 97295087.  
 RA MAYNARD J., KRANCZAK M., UPADHYAYA M.;  
 RL HUM. GENET. 99:674-676(1997).  
 RN [22]  
 RP VARIANT NF1 ARG-1952.  
 RX MEDLINE: 97255969.  
 RA HUDSON J., WU C.L., TASSABEHI M., SUMMERS E.M., SIMON S., SUPPER M.,  
 RA DONNAI D., THAKKER N.;  
 RL HUM. MUTAT. 9:366-367(1997).  
 RN [23]  
 RP VARIANT NF1 TRP-1611.  
 RX MEDLINE: 97442280.  
 RA UPADHYAYA M., MAYNARD J., OSBORN M., HARPER P.S.;  
 RL HUM. MUTAT. 10:248-250(1997).  
 CC -1- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NF1 SHOWS GREATER  
 CC AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE  
 CC A REGULATOR OF RAS ACTIVITY.  
 CC -1- DISEASE: THIS PROTEIN IS ASSOCIATED WITH TYPE 1 NEUROFIBROMATOSIS  
 CC (NF1) (ALSO CALLED VON RECKLINGHAUSEN SYNDROME), THE MOST FREQUENT

CC INHERITED GENETIC DISEASE (ABOUT 1 IN 3000). IT EXHIBITS FULL  
 CC PENETRANCE AND HIGH MUTATION RATE WITH 30 TO 50% OF NF1 PATIENTS  
 CC REPRESENTING A NEW MUTATION. AMONG THE MANY CLINICAL FEATURES OF  
 CC NF1 ARE PATCHES OF SKIN PIGMENTATION (CAFE-AU-LAIT SPOTS), LISCH  
 CC NODULES OF THE IRIS PERIPHERAL, PERIPHERAL NERVOUS SYSTEM  
 CC ASSOCIATED TUMORS AND FIBROMATOUS SKIN TUMORS. THE DISEASE  
 CC DEMONSTRATES A HIGH DEGREE OF PENETRANCE BY AGE 5 YEARS.  
 CC -1- DISEASE: DEFECTS IN NF1 ARE ASSOCIATED WITH WATSON SYNDROME (WS).  
 CC WS IS CHARACTERIZED BY THE PRESENCE OF PULMONARY STENOSIS,  
 CC CAFE-AU-LAIT SPOTS, AND MENTAL RETARDATION. WS IS CONSIDERED AS  
 CC AN ATYPICAL FORM OF NF1.  
 CC -1- DISEASE: DEFECTS IN NF1 ARE ASSOCIATED WITH LEOPARD, AN AUTOSOMAL  
 CC DOMINANT SYNDROME. LEOPARD (LS) IS AN ACROPHORY FOR THE FEATURES OF  
 CC THIS SYNDROME: L-LENTIGINES ("FRECKLES"), P-ELECTROCARDIOGRAPHIC  
 CC ABNORMALITIES, O-OCULAR HYPERTELOPSIS, P-PULMONARY STENOSIS,  
 CC A-ABNORMALITIES OF GENITALIA, R-RETARDATION OF GROWTH, AND  
 CC D-DEAFNESS (SENSORY-NEURAL). THE MAIN FEATURES OF THE SYNDROME ARE  
 CC MULTIPLE LENTIGINES IN COMBINATION WITH A CONGENITAL HEART  
 CC MALFORMATION (PULMONARY STENOSIS, SUBVALVULAR MUSCULAR AORTIC  
 CC STENOSIS). A CLINICAL OVERLAP EXISTS BETWEEN LS, NF1 AND WS.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF NEUROFIBROMIN ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING OF THE NF1 GENE: TYPE I AND TYPE II (SHOWN  
 CC HERE) ONLY DIFFERS IN THE INSERTION OF A 21 RESIDUES SEGMENT.  
 CC -1- SIMILARITY: TO OTHER RAS GTPASE-ACTIVATING PROTEINS.  
 DR EMBL: M82814; G189165; -;  
 DR EMBL: M89914; G292354; -;  
 DR EMBL: M8116; G494225; -;  
 DR EMBL: M38108; G494225; JOINED.  
 DR EMBL: M38109; G494225; JOINED.  
 DR EMBL: M38110; G494225; JOINED.  
 DR EMBL: M38111; G494225; JOINED.  
 DR EMBL: M38112; G494225; JOINED.  
 DR EMBL: M38113; G494225; JOINED.  
 DR EMBL: M38114; G494225; JOINED.  
 DR EMBL: M38106; G189170; -;  
 DR EMBL: M61213; G189163; -;  
 DR EMBL: M38107; G189172; ALT\_SEQ.  
 DR EMBL: M60496; G189158; -;  
 DR EMBL: M60915; G189161; -;  
 DR EMBL: S51751; G262288; -;  
 DR EMBL: D12625; G219940; ALT\_SEQ.  
 DR PIR: A35222; A35222.  
 DR PIR: A35605; A35605.  
 DR PIR: A35879; A35879.  
 DR PIR: J01277; J01277.  
 DR MIM: 162200; -;  
 DR MIM: 193520; -;  
 DR MIM: 151100; -;  
 DR PROSITE: PS00509; RAS\_GTPASE\_ACTIV\_1; 1.  
 DR PROSITE: PS50018; RAS\_GTPASE\_ACTIV\_2; 1.  
 KW GTPASE ACTIVATION; ALTERNATIVE SPLICING; ANTI-ONCOGENE;  
 Note: remainder of annotations omitted.  
 Query Match 72.0% Score 54; DB 1; Length 2839;  
 Best Local Similarity 55.6% Pred. No. 3.68e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Db 2156 FSLPKFYL 2164  
 Qy 1 FAMNPFYL 9  
 RESULT 7  
 ID NF1\_MOUSE STANDARD; PRT; 2841 AA.  
 AC 004690; 061956; 061957;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).  
 GN NF1.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:

OC EUTHERIA; RODENTIA.  
 RN (1)  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-BALB/C; TISSUE-BRAIN;  
 RX MEDLINE: 93357730.  
 RA BERNARDS A., SNIDJERS A.J., HANNIGAN G.E., MURPHY A.E., GUSELLA J.F.,  
 RL HOM. MOL. GENET. 2:645-650(1993).  
 RN (2)  
 RN SEQUENCE OF 1178-1555 FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE: 95047432.  
 RA MANIANT A., MANASUGI S., YOKOTA Y., ABE K., USHIO Y., YAMAMURA K.,  
 RL GENE 148:245-251(1994).  
 RN (3)  
 RN SEQUENCE OF 1950-2568 FROM N.A.  
 RX MEDLINE: 90384569.  
 RA BUCHBERG A.M., CLEVELAND L.S., JENKINS N.A., COPELAND N.G.,  
 RL NATURE 347:291-294(1990).  
 CC -1- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NFI SHOWS GREATER  
 CC AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE  
 CC A REGULATOR OF RAS ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: TYPE I IS EXPRESSED PREDOMINANTLY IN BRAIN,  
 CC SPINAL CORD AND TESTIS. TYPE II IS EXPRESSED PREDOMINANTLY IN  
 CC ADRENAL GLAND, KIDNEY, OVARY AND LUNG. TYPE III IS EXPRESSED  
 CC PREDOMINANTLY IN ADRENAL GLAND AND TYPE IV IS EXPRESSED  
 CC MAINLY IN THE TESTIS.  
 CC -1- ALTERNATIVE PRODUCTS: FOUR FORMS OF THE PROTEIN (TYPES I, II,  
 CC III AND IV) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME  
 CC GENE. THE SEQUENCE SHOW HERE IS THAT OF TYPE II.  
 CC -1- SIMILARITY: TO OTHER RAS GTPASE-ACTIVATING PROTEINS.  
 DR EMBL: L10369; G309451; -  
 DR EMBL: L10367; G309451; JOINED.  
 DR EMBL: L10368; G309451; JOINED.  
 DR EMBL: L10370; G309453; -  
 DR EMBL: X54924; G930191; -  
 DR EMBL: D30730; G577638; -  
 DR EMBL: D30731; G577640; -  
 DR MGD: MGI:97306; NFI.  
 DR PROSITE: PS00509; RAS\_GTPASE\_ACTIV\_1; 1.  
 DR PROSITE: PS0018; RAS\_GTPASE\_ACTIV\_2; 1.  
 KW GTPASE ACTIVATION; ALTERNATIVE SPLICING.  
 FT DOMAIN 1237 1453 RAS-GAP.  
 FT VARSPLIC 1373 1393 MISSING (IN TYPE I AND TYPE IV).  
 FT VARSPLIC 1394 1406 VYSORPONSIGA -> VPKSCSCINRWLASLRT  
 FT VARSPLIC 1407 2841 ASVP (IN TYPE III AND TYPE IV).  
 FT VARSPLIC 2841 319591 MISSING (IN TYPE III AND TYPE IV).  
 FT VARSPLIC 319591 MM: A7AA/6F4 CRC32;  
 SQ SEQUENCE 2841 AA: 319591 MW: 319591 MW: A7AA/6F4 CRC32;  
 Query Match 72.0%; Score 54; DB 1; Length 2841;  
 Best Local Similarity 55.6%; Pred. No. 3.68e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Db 2158 FSLPKFYLL 2166  
 Oy 1 FAMPNFTL 9  
 RESULT 8  
 ID YN8S\_YEAST STANDARD; PRT; 393 AA.  
 AC P53740;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 44.5 KD PROTEIN IN PELT494-MSOI INTERGENIC REGION.  
 GN YNR048W OR N3453.  
 OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA FOHL T.M.;  
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- SIMILARITY: TO YEAST YCR94W AND YNL323W.  
 DR EMBL: Z71663; E239593; -  
 KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.

FT TRANSMEM 47 67 POTENTIAL.  
 FT TRANSMEM 335 355 POTENTIAL.  
 SQ SEQUENCE 393 AA: 44542 MW: 4660346A CRC32;  
 Query Match 69.3%; Score 52; DB 1; Length 393;  
 Best Local Similarity 75.0%; Pred. No. 8.63e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Db 280 ALPNFTKL 287  
 Oy 2 AMPNFTL 9  
 RESULT 9  
 ID TYRA\_LACLA STANDARD; PRT; 354 AA.  
 AC P43901;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH).  
 GN TYRA.  
 OS LACTOCOCCUS LACTIS (SUBSP. LACTIS) (STREPTOCOCCUS LACTIS).  
 OC PROKARYOTA; FIRMICUTES; COCCI; STREPTOCOCCAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-F15876;  
 RX MEDLINE: 95124293.  
 RA GRIFFIN H.G., GASSON M.J.;  
 RL MOL. GEN. GENET. 246:119-127(1995).  
 CC -1- CATALYTIC ACTIVITY: PREPHENATE + NAD(+) -> 4-HYDROXYPHENILPYRUVATE  
 CC + CO(2) + NADH.  
 CC -1- PATHWAY: TYROSINE BIOSYNTHESIS.  
 DR EMBL: X78413; G683582; -  
 KW TYROSINE BIOSYNTHESIS; OXIDOREDUCTASE; NAD.  
 FT NP\_BIND 3 33 NAD (POTENTIAL).  
 FT NP\_BIND 354 AA: 39530 MW: DD0FE758 CRC32;  
 SQ SEQUENCE 354 AA: 39530 MW: DD0FE758 CRC32;  
 Query Match 68.0%; Score 51; DB 1; Length 354;  
 Best Local Similarity 75.0%; Pred. No. 1.31e+01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Db 281 ALPNFTL 288  
 Oy 2 AMPNFTL 9  
 RESULT 10  
 ID AMPN\_ECOLI STANDARD; PRT; 869 AA.  
 AC P04825;  
 DT 13-AUG-1987 (REL. 05, CREATED)  
 DT 13-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE AMINOPEPTIDASE N (EC 3.4.11.2) (ALPHA-AMINOACYLPEPTIDE HYDROLASE).  
 GN PERP.  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA FOGLINO M., GHARBI S., LAZDUNSKI A.;  
 RL GENE 49:303-309(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-K12;  
 RX MEDLINE: 87193509.  
 RA MCCAMAN M.T., GABE J.D.;  
 RL GENE 48:145-153(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA BLATTNER F.R., BLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

[4] SEQUENCE FROM N.A.  
 RN STRAIN-K12;  
 RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,  
 RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K.,  
 RA KASAI H., KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M.,  
 RA KITAKAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K., MORI H.,  
 RA MOTOMODA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T.,  
 RA SATO N., SAMEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C.,  
 RA YANAMOTO Y., YANO M.,  
 RN SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN (5)  
 RP SEQUENCE OF 1-241 FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE: 86310300.  
 RA MCCAMMAN M.T., GABE J.D.;  
 RL MOL. GEN. GENET. 204:148-152(1986).  
 RN (6)  
 RP SEQUENCE OF 1-176 FROM N.A., AND SEQUENCE OF 1-21.  
 RX MEDLINE: 86164315.  
 RA BALLY M., FOGIINO M., BRUSCHI M., MURGIER M., LAZDUNSKI A.;  
 RL EUR. J. BIOCHEM. 155:565-569(1986).  
 CC -1- FUNCTION: AMINOPEPTIDASE N IS INVOLVED IN THE DEGRADATION OF  
 CC INTRACELLULAR PEPTIDES GENERATED BY PROTEIN BREAKDOWN DURING  
 CC NORMAL GROWTH AS WELL AS IN RESPONSE TO NUTRIENT STARVATION.  
 CC -1- COFACTOR: BINDS AND REQUIRES A ZINC ATOM, WHICH IS ESSENTIAL FOR  
 CC PROTEOLYTIC ACTIVITY.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, BOUND TO THE INNER FACE OF  
 CC THE CYTOPLASMIC MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1 (ZINC METALLOPROTEASES);  
 CC ALSO KNOWN AS THE PEPN SUBFAMILY.  
 DR EMBL: X04020: G42353;  
 DR EMBL: X03709: G42356;  
 DR EMBL: M15676: G147144;  
 DR EMBL: AE000195: G1787163;  
 DR EMBL: D90731: G1651457;  
 DR EMBL: D90732: G1651459;  
 DR EMBL: M15273: G147142;  
 DR PIR: A29045: DPECN.  
 DR PIR: A27164: A27164.  
 DR ECOGENE: EG10696: PEPN.  
 DR PROSITE: PS00142: ZINC\_PROTEASE; 1.  
 KW HYDROLASE; METALLOPROTEASE; AMINOPEPTIDASE; ZINC; MEMBRANE.  
 FT METAL 0  
 FT INIT MET 0  
 FT ACT\_SITE 296 296 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 297 297 BY SIMILARITY.  
 FT METAL 300 300 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 319 319 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 380 380 PROTON DONOR (POTENTIAL).  
 FT CONFLICT 75 75 E -> D (IN REF. 4 AND 6).  
 SO SEQUENCE 869 AA; 98787 MW; ADA0286A CRC32;  
 Query Match 68.0%; Score 51; DB 1; Length 869;  
 Best Local Similarity 66.7%; Pred. No. 1.31e+01;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 369 IEMNFYTL 377  
 Oy 1 FAMPNFYTL 9  
 RESULT 11  
 ID NUGC MARPO STANDARD; PRT: 169 AA.  
 AC P12199;  
 DT 01-OCT-1989 (REL. 12, CREATED)  
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE NADH-PLASTOQUINONE OXIDOREDUCTASE SUBUNIT J (EC 1.6.5.3) (ORF 169).  
 GN NADH  
 OS MARCHANTIA POLYMORPHA (LIVERMORT).  
 OG CHLOROPLAST.  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; BRYOPHYTA; HEPATICOPSIDA.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA OHYAMA K.;  
 RL SUBMITTED (OCT-1986) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN (2)  
 RP COMPLETE GENOME.  
 RA OHYAMA K., FUKUYAMA H., KOHCHI T., SHIRAI H., SANO T., SANO S.,  
 RA UMEZONO K., SHIKI Y., TAKEUCHI M., CHANG Z., AOTA S., INOKUCHI H.,  
 RA OZAKI H.;  
 RL NATURE 322:572-574(1986).  
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE - NAD(+) + UBIQUINOL.  
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 30 KD SUBUNIT FAMILY.  
 DR EMBL: X04465: G11675;  
 DR PIR: A05042: A05042;  
 DR PIR: S01601: S01601.  
 DR PROSITE: PS00542: COMPLEX\_I\_30K; 1.  
 KW OXIDOREDUCTASE; NAD; PLASTOQUINONE; CHLOROPLAST.  
 SO SEQUENCE 169 AA; 20085 MW; 8BC10865 CRC32;  
 Query Match 66.7%; Score 50; DB 1; Length 169;  
 Best Local Similarity 55.6%; Pred. No. 1.97e+01;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Db 157 YVNFYEL 165  
 Oy 1 FAMPNFYTL 9  
 RESULT 12  
 ID PENT\_RAT STANDARD; PRT: 198 AA.  
 AC Q08388;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE PHOSPHATIDYLETHANOLAMINE N-METHYLTRANSFERASE (EC 2.1.1.17) (PENT).  
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 93346418.  
 RA CUI Z., VANCE J.E., CHEN M.H., VOELKER D.R., VANCE D.E.;  
 RL J. BIOL. CHEM. 268:16655-16663(1993).  
 RN [2]  
 RP SEQUENCE OF 1-30.  
 RA RIDGWAY N.D.;  
 RL THEISIS (1988), UNIVERSITY OF BRITISH COLUMBIA, CANADA.  
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + PHOSPHATIDYL-  
 CC ETHANOLAMINE - S-ADENOSYL-L-HOMOCYSTEINE + PHOSPHATIDYL-N-  
 CC METHYLETHANOLAMINE.  
 CC -1- PATHWAY: FIRST, SECOND AND THIRD STEPS OF PHOSPHATIDYLETHANOLAMINE  
 CC METHYLATION PATHWAY.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: LIVER.  
 CC -1- SIMILARITY: TO YEAST PEM2.  
 DR EMBL: L14441: G310195;  
 KW PHOSPHOLIPID BIOSYNTHESIS; TRANSFERASE; METHYLTRANSFERASE;  
 KW TRANSMEMBRANE.  
 FT INIT\_MET 0  
 FT TRANSMEM 12 32 POTENTIAL.  
 FT TRANSMEM 45 65 POTENTIAL.  
 FT TRANSMEM 90 110 POTENTIAL.  
 FT TRANSMEM 158 178 POTENTIAL.  
 SO SEQUENCE 198 AA; 22355 MW; 90AED9AB CRC32;  
 Query Match 66.7%; Score 50; DB 1; Length 198;  
 Best Local Similarity 44.4%; Pred. No. 1.97e+01;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 Db 102 FVLSFYAL 110  
 Oy 1 FAMPNFYTL 9



## RESULT 13

ID SP51.BACSU STANDARD; PRT: 246 AA.  
 AC P39629;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE SPORE COAT POLYSACCHARIDE BIOSYNTHESIS PROTEIN SPSI.  
 GN SPSI OR IPA-71D  
 OS BACILLUS SUBTILIS.  
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE: 95020537.  
 RA GLASER P., KUNST F., ARNAUD M., COUDART M.P., GONZALES W.,  
 RA HOLLO M.F., IONESCU M., LUBOCHINSKY B., MARCELINO L., MOSZER I.,  
 RA PRESCAN E., SANTANA M., SCHNEIDER E., SCHWEIZER J., VERTES A.,  
 RA RAPPORT G., DANCHIN A.;  
 RL MOL. MICROBIOL. 10:371-384(1993).  
 CC -1- PATHWAY: SPORE COAT POLYSACCHARIDE BIOSYNTHESIS.  
 CC -1- SIMILARITY: TO GLUCOSE-1-PHOSPHATE THYMIDYLKINASES.  
 DR EMBL: X3124; G580880;  
 DR SUBTILIST: BG10617; SPSI.  
 KW TRANSFERASE: KINASE; NUCLEOTIDYLTRANSFERASE.  
 SQ SEQUENCE 246 AA; 27773 MW; 921EF443 CRC32;

## Query Match

Best Local Similarity 71.4%; Score 50; DB 1; Length 246;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 58 MPOFYKL 64  
 |||||  
 QY 3 MPNFTYL 9

## RESULT 14

ID Y926.HELRY STANDARD; PRT: 381 AA.  
 AC P55985;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL PROTEIN HP0926.  
 GN HP0926.  
 OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;  
 OC AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695;  
 RX MEDLINE: 97394467.  
 RA TOMB J.F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,  
 RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,  
 RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,  
 RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODER A.,  
 RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKET E.K.,  
 RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,  
 RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATHEY L., WALLIN E.,  
 RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,  
 RA VENTER J.C.;  
 RL NATURE 388:539-547(1997).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0024 FAMILY.  
 CC PROSITE: PS01268; UPF0024; 1.  
 DR TIGR: HP0926;  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 381 AA; 44003 MW; C659A962 CRC32;

## Query Match

Best Local Similarity 66.7%; Score 50; DB 1; Length 381;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 159 FGMPNPF 165  
 |||||

## QY 1 FAMPNFY 7

RESULT 15  
 ID YE28.CAEEL STANDARD; PRT: 410 AA.  
 AC Q20585;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 47.6 KD PROTEIN F49C12.8 IN CHROMOSOME IV.  
 GN F49C12.8.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA GARDNER A.;  
 RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- SIMILARITY: BELONGS TO THE FUS6 FAMILY.  
 DR EMBL: Z68227; E214023;  
 DR WORMPEP: F49C12.8; CE03368.  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 410 AA; 47583 MW; 44CFB5F2 CRC32;

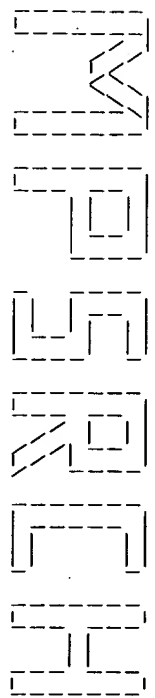
## Query Match

Best Local Similarity 66.7%; Score 50; DB 1; Length 410;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 62 FDMAPFYL 70  
 |||||  
 QY 1 FAMPNFTYL 9

Search completed: Fri Sep 11 12:59:13 1998  
 Job time : 7 secs.

**THIS PAGE BLANK (USPTO)**



Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Sep 11 12:59:31 1998; MasPar time 3.57 Seconds  
Tabular output not generated. 106.103 Million cell updates/sec

Title: >US-08-452-843-6  
Description: (1-9) from US08452843.pep  
Perfect Score: 75  
Sequence: 1 FAMPNFTTL 9

Scoring table:  
Gap 15  
PAM 150

Searched: 140555 seqs, 42109429 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: 1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mhc 6:sp\_organelle 7:sp\_phage 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp\_vertebrate  
13:sp\_unclassified

Statistics: Mean 24.487; Variance 33.512; scale 0.731  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description            | Pred. No. |
|------------|-------|-------------|--------|----|--------|------------------------|-----------|
| 1          | 56    | 74.7        | 319    | 9  | 029900 | C4-DICARBOXYLATE TRANS | 2.56e+00  |
| 2          | 55    | 73.3        | 418    | 9  | 022625 | T21B10.4.              | 3.97e+00  |
| 3          | 55    | 73.3        | 786    | 9  | 031874 | YOSO PROTEIN.          | 3.97e+00  |
| 4          | 54    | 72.0        | 189    | 9  | 026218 | HYPOTHETICAL 21.7 KD P | 6.12e+00  |
| 5          | 54    | 72.0        | 287    | 9  | 005943 | CYTOCHROME OXIDASE D,  | 6.12e+00  |
| 6          | 54    | 72.0        | 466    | 3  | 017451 | GAG-LIKE PROTEIN.      | 6.12e+00  |
| 7          | 54    | 72.0        | 2764   | 3  | 001399 | NEUROFIBROMIN.         | 6.12e+00  |
| 8          | 54    | 72.0        | 2802   | 3  | 001397 | NEUROFIBROMIN.         | 6.12e+00  |
| 9          | 54    | 72.0        | 2802   | 3  | 001398 | NEUROFIBROMIN.         | 6.12e+00  |
| 10         | 54    | 72.0        | 2820   | 10 | P97526 | NEUROFIBROMIN.         | 6.12e+00  |
| 11         | 53    | 70.7        | 724    | 3  | 023874 | D2 ORF (FRAGMENT).     | 9.39e+00  |
| 12         | 53    | 69.3        | 743    | 3  | 023869 | D2 ORF (FRAGMENT).     | 9.39e+00  |
| 13         | 52    | 69.3        | 260    | 3  | 059032 | D2 ORF (FRAGMENT).     | 1.43e+01  |
| 14         | 52    | 69.3        | 275    | 3  | 017711 | C55A1.1.               | 1.43e+01  |
| 15         | 52    | 69.3        | 341    | 3  | 018101 | T21B4.5.               | 1.43e+01  |
| 16         | 52    | 69.3        | 364    | 3  | P91384 | COSMID K12D9.          | 1.43e+01  |
| 17         | 52    | 69.3        | 426    | 8  | 023842 | S GLYCOPROTEIN (FRAGME | 1.43e+01  |
| 18         | 52    | 69.3        | 429    | 8  | 023845 | S GLYCOPROTEIN (FRAGME | 1.43e+01  |
| 19         | 52    | 69.3        | 931    | 11 | P87544 | 104K PROTEIN.          | 1.43e+01  |
| 20         | 51    | 68.0        | 307    | 9  | P95159 | HYPOTHETICAL 33.2 KD P | 2.17e+01  |

|    |    |      |      |    |        |                        |          |
|----|----|------|------|----|--------|------------------------|----------|
| 21 | 51 | 68.0 | 605  | 9  | P72607 | ABC TRANSPORTER.       | 2.17e+01 |
| 22 | 51 | 68.0 | 1464 | 11 | 066951 | E2 GLYCOPROTEIN PRECUR | 2.17e+01 |
| 23 | 50 | 66.7 | 199  | 10 | 061907 | PHOSPHATIDYLETHANOLAMI | 3.27e+01 |
| 24 | 50 | 66.7 | 306  | 3  | 020457 | F46C3.2.               | 3.27e+01 |
| 25 | 50 | 66.7 | 343  | 3  | 019572 | SIMILAR TO GUANINE NUC | 3.27e+01 |
| 26 | 50 | 66.7 | 502  | 3  | 021291 | K07F5.6.               | 3.27e+01 |
| 27 | 50 | 66.7 | 1001 | 3  | 001261 | T2003.9.               | 3.27e+01 |
| 28 | 49 | 65.3 | 334  | 3  | 017970 | C14C10.1.              | 4.90e+01 |
| 29 | 49 | 65.3 | 409  | 8  | 039363 | S-LOCUS GLYCOPROTEIN P | 4.90e+01 |
| 30 | 49 | 65.3 | 429  | 8  | 023843 | S GLYCOPROTEIN (FRAGME | 4.90e+01 |
| 31 | 49 | 65.3 | 429  | 8  | 023838 | S GLYCOPROTEIN (FRAGME | 4.90e+01 |
| 32 | 49 | 65.3 | 431  | 8  | 023849 | S GLYCOPROTEIN (FRAGME | 4.90e+01 |
| 33 | 49 | 65.3 | 431  | 8  | 023861 | S GLYCOPROTEIN (FRAGME | 4.90e+01 |
| 34 | 49 | 65.3 | 431  | 8  | 023852 | S GLYCOPROTEIN (FRAGME | 4.90e+01 |
| 35 | 49 | 65.3 | 436  | 8  | 041222 | S GLYCOPROTEIN (FRAGME | 4.90e+01 |
| 36 | 49 | 65.3 | 437  | 8  | 039278 | SLG12 (FRAGMENT).      | 4.90e+01 |
| 37 | 49 | 65.3 | 749  | 3  | 017950 | K12G11.1.              | 4.90e+01 |
| 38 | 49 | 65.3 | 857  | 8  | 039393 | S-RECEPTOR KINASE-LIKE | 4.90e+01 |
| 39 | 49 | 65.3 | 858  | 8  | 001963 | SERINE/THREONINE KINAS | 4.90e+01 |
| 40 | 49 | 65.3 | 952  | 11 | 036413 | TEGMENT PROTEIN.       | 4.90e+01 |
| 41 | 48 | 64.0 | 243  | 9  | 045943 | ORF 243.               | 7.30e+01 |
| 42 | 48 | 64.0 | 549  | 9  | 029570 | LONG-CHAIN-FATTY-ACID- | 7.30e+01 |
| 43 | 48 | 64.0 | 1274 | 1  | 006673 | CHROMOSOME XII COSMID  | 7.30e+01 |
| 44 | 48 | 64.0 | 3023 | 11 | 084898 | GENOME POLYPROTEIN (CO | 7.30e+01 |
| 45 | 48 | 64.0 | 3023 | 11 | 088925 | POLYPROTEIN.           | 7.30e+01 |

ALIGNMENTS

RESULT 1  
ID 029900  
AC 029900; PRELIMINARY; PRT; 319 AA.  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE C4-DICARBOXYLATE TRANSPORTER (MAL1).  
GN AF0347.  
OS ARCHAEoglobus fulgidus.  
OC ARCHAEABACTERIA; EURYARCHAEOTA; ARCHAEoglobales; ARCHAEoglobaceae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KLEIN H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E., KETCHUM K.A.,  
RA DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L.,  
RA KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C., FLEISCHMANN R.D.,  
RA OUCKENBUSH J., LEE N.H., SUTTON G.G., GILL S., KIRKNESS E.F.,  
RA DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B., PETERSON S.,  
RA REICH C.I., MCNEIL L.K., BADGER J.H., GLOCKER A., ZHOU L., OVERBECK R.,  
RA GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T., COTTON M.D.,  
RA SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M., SADOW P.W.,  
RA D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M.,  
RA OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.,  
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA KLEIN H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E., KETCHUM K.A.,  
RA DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L.,  
RA KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C., FLEISCHMANN R.D.,  
RA OUCKENBUSH J., LEE N.H., SUTTON G.G., GILL S., KIRKNESS E.F.,  
RA DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B., PETERSON S.,  
RA REICH C.I., MCNEIL L.K., BADGER J.H., GLOCKER A., ZHOU L., OVERBECK R.,  
RA GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T., COTTON M.D.,  
RA SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M., SADOW P.W.,  
RA D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M.,  
RA OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.,  
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL, AB001080; G250284; ..  
SQ SEQUENCE 319 AA; 35267 MW; 6F13B082 CRC32;  
Query Match 74.7%; Score 56; DB 9; Length 319;  
Best Local Similarity 66.7%; Pred. No. 2.56e+00;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
DB 68 FVGNFTPL 76



DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
 DE CYTOCHROME OXIDASE D, SUBUNIT II (FRAGMENT).  
 GN CYDB.  
 OS RICKETTSIA PROMAZEEKII.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; RICKETTSIAS; RICKETTSIALDS;  
 OC RICKETTSINCEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E.  
 RA ANDERSSON J.O.; ANDERSSON S.G.E.;  
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: Y11780; E307800;  
 FT NON\_TER 1  
 SQ SEQUENCE 287 AA; 32395 MW; C19BAA80 CRC32;  
 Query Match 72.0%; Score 54; DB 9; Length 287;  
 Best Local Similarity 66.7%; Pred. No. 6.12e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 DB 178 FSLPREFYL 186  
 QY 1 FAMPNFTYL 9  
 RESULT 6  
 ID 017451 PRELIMINARY; PRT; 466 AA.  
 AC 017451;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE GAG-LIKE PROTEIN.  
 GN GAG.  
 OS CULEX PIPIENS (HOUSE MOSQUITO).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BENSADT-MERCHERER N.; CAGNON C.; DESMONS I.; SALVADO J.C.; KARANA S.,  
 RA D'AMICO F.; MOUTCHES C.;  
 RL GENETICA 0:0-0(1997).  
 DR EMBL: AF030588; G2623228;  
 SQ SEQUENCE 466 AA; 51269 MW; 270BA437 CRC32;  
 Query Match 72.0%; Score 54; DB 3; Length 466;  
 Best Local Similarity 44.4%; Pred. No. 6.12e+00;  
 Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 DB 426 FTLPEFYL 434  
 QY 1 FAMPNFTYL 9  
 RESULT 7  
 ID 001399 PRELIMINARY; PRT; 2764 AA.  
 AC 001399;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
 DE NEUROFIBROMIN.  
 GN NFL.  
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON S;  
 RA HANNIGAN G.E.; THE I.; SHAMANSKI F.L.; ORR-WEAVER T.L.; GUSELLA J.F.,  
 RA BERNARDS A.;  
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: L26502; G1929433;  
 DR FLYBASE: FBgn0015269; NFL.  
 SQ SEQUENCE 2764 AA; 312934 MW; 54B6B40F CRC32;  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

Query Match 72.0%; Score 54; DB 3; Length 2764;  
 Best Local Similarity 55.6%; Pred. No. 6.12e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 DB 2187 FSLPREFYL 2195  
 QY 1 FAMPNFTYL 9  
 RESULT 8  
 ID 001397 PRELIMINARY; PRT; 2802 AA.  
 AC 001397;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
 DE NEUROFIBROMIN.  
 GN NFL.  
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON S;  
 RA HANNIGAN G.E.; THE I.; SHAMANSKI F.L.; ORR-WEAVER T.L.; GUSELLA J.F.,  
 RA BERNARDS A.;  
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: L26500; G1929429;  
 DR FLYBASE: FBgn0015269; NFL.  
 SQ SEQUENCE 2802 AA; 317202 MW; 032CE079 CRC32;  
 Query Match 72.0%; Score 54; DB 3; Length 2802;  
 Best Local Similarity 55.6%; Pred. No. 6.12e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 DB 2187 FSLPREFYL 2195  
 QY 1 FAMPNFTYL 9  
 RESULT 9  
 ID 001398 PRELIMINARY; PRT; 2802 AA.  
 AC 001398;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
 DE NEUROFIBROMIN.  
 GN NFL.  
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON S;  
 RA HANNIGAN G.E.; THE I.; SHAMANSKI F.L.; ORR-WEAVER T.L.; GUSELLA J.F.,  
 RA BERNARDS A.;  
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: L26501; G1929431;  
 DR FLYBASE: FBgn0015269; NFL.  
 SQ SEQUENCE 2802 AA; 317209 MW; 76822162 CRC32;  
 Query Match 72.0%; Score 54; DB 3; Length 2802;  
 Best Local Similarity 55.6%; Pred. No. 6.12e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 DB 2187 FSLPREFYL 2195  
 QY 1 FAMPNFTYL 9  
 RESULT 10  
 ID P97526 PRELIMINARY; PRT; 2820 AA.  
 AC P97526;  
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE NEUROFIBROMIN.  
GN NFI.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MISTAR; TISSUE-BRAIN;  
RA KYRITSIS A.P., LEE P.S., MOCHIZUKI H., NISHI T., LEVIN V.A., SAYA H.;  
RL INT. J. ONCOL. 1:149-152(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MISTAR; TISSUE-BRAIN;  
RA SUZUKI H., TAKAHASHI K., YASUMOTO K.I., FUSE N., SHIBAHARA S.;  
RL J. BIOCHEM. 120:1048-1054(1996).  
DR EMBL; D45201; G1841314;  
DR PROSITE; PS00509; RAS\_GTPASE\_ACTIV.1; 1.  
SQ SEQUENCE 2820 AA; 317079 MW; 64708267 CRC32;

Query Match 72.0%; Score 54; DB 10; Length 2820;  
Best Local Similarity 55.6%; Pred. No. 6.12e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2137 FSLPKFYL 2145  
1 FAMPNFYTL 9

RESULT 11  
ID Q23874 PRELIMINARY; PRT; 724 AA.  
AC Q23874;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE D2 ORF (FRAGMENT);  
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).  
OC EUKARYOTA; PROTOZOA; SARCOMASTICOPHORA; SARCODINA; RHIZOPODA;  
OC EUMYCETOZOA; BICTYOSTELIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NC4;  
RA KIYOSAWA H., HUGHES J.E., WELKER D.L.;  
RL SUBMITTED (NOV-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; U00796; G392877;  
FT NON-TER 724  
SQ SEQUENCE 724 AA; 82631 MW; 2FC1BD50 CRC32;

Query Match 70.7%; Score 53; DB 3; Length 724;  
Best Local Similarity 55.6%; Pred. No. 9.39e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 698 FEMPNLESL 706  
1 FAMPNFYTL 9

RESULT 12  
ID Q23869 PRELIMINARY; PRT; 743 AA.  
AC Q23869;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE D2 ORF.  
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).  
OC EUKARYOTA; PROTOZOA; SARCOMASTICOPHORA; SARCODINA; RHIZOPODA;  
OC EUMYCETOZOA; DICTYOSTELIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NC4;  
RA MEDLINE; 94302138.  
RA FARRAR N.A., KIYOSAWA H., HUGHES J.E., WELKER D.L., WILLIAMS K.L.;  
RL PLASMID 31:184-195(1994).

DR EMBL; U00691; G392796;  
KM PLASMID.  
SQ SEQUENCE 743 AA; 84863 MW; B5E80804 CRC32;

Query Match 70.7%; Score 53; DB 3; Length 743;  
Best Local Similarity 55.6%; Pred. No. 9.39e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 717 FEMPNLESL 725  
1 FAMPNFYTL 9

RESULT 13  
ID Q59032 PRELIMINARY; PRT; 260 AA.  
AC Q59032;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE HYPOTHETICAL 30.0 KD PROTEIN 1638.  
GN M31638.  
OS METHANOCOCCUS JANNASCHII.  
OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BUT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A.,  
RA GOCAYNE J.D., KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.F.,  
RA ADAMS M.D., REICH C.I., OVERBEK R., KIRNESS E.F.,  
RA WEINSTOCK K.G., MERRICK J.M., GLODER A., SCOTT J.L.,  
RA GEORGAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., PRESLEY E.A.,  
RA NGUYEN D., UTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W.,  
RA HANNA M.C., COTTON M.D., HORST W.A., ROBERTS K.M., KAINE B.P.,  
RA BORODOVSKY M., KLEIN H.P., FRASER C.M., SMITH H.O., WOESE C.R.,  
RA VENTER J.C.;  
RL SCIENCE 273:1058-1073(1996).  
DR EMBL; U67604; G1500538;  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 260 AA; 30016 MW; 46BB0D51 CRC32;

Query Match 69.3%; Score 52; DB 9; Length 260;  
Best Local Similarity 44.4%; Pred. No. 1.43e+01;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 185 FNLSPFAL 193  
1 FAMPNFYTL 9

RESULT 14  
ID 017711 PRELIMINARY; PRT; 275 AA.  
AC 017711;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE C55A1.1.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAWKINS T., HILLER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
RA MCMURRAY A., MORITMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIFKEN L., ROOPRA A., SANDERS D., SHOWKNEEN R., SMALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., STULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 368:32-38(0).  
DR EMBL; Z81489; E348489;  
SQ SEQUENCE 275 AA; 31417 MW; 8C8DC89E CRC32;

Sun, Sep 13 10:57:04 1998

Query Match 69.3%; Score 52; DB 3; Length 275;  
 Best Local Similarity 66.7%; Pred. No. 1.43e+01;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 108 FAFPIFYAL 116  
 OY 1 FAMPNFYTL 9

RESULT 15  
 ID 018101 PRELIMINARY; PRT; 341 AA.  
 AC 018101;  
 DT 01-JAN-1998 (TREMREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMREL. 05, LAST ANNOTATION UPDATE)  
 DE T21B4.5.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SMYE R.;  
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERRS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SWALDON N., SMITH A., SONNHAMER E., STADEN R., SULSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
 RL NATURE 368:32-38(0).  
 DR EMBL; Z81124; E1180983;  
 SO SEQUENCE 341 AA; 38632 MW; 3D79791E CRC32;

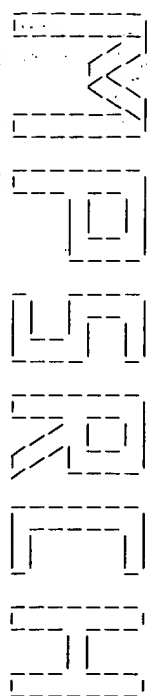
Query Match 69.3%; Score 52; DB 3; Length 341;  
 Best Local Similarity 71.4%; Pred. No. 1.43e+01;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 82 FAMPSEFY 88  
 OY 1 FAMPNFY 7

Search completed: Fri Sep 11 13:00:02 1998  
 Job time : 31 secs.

**THIS PAGE BLANK (USPTO)**





(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 12:53:33 1998; Maspar time 2.54 Seconds

Tabular output not generated. 57,346 Million cell updates/sec

Title: >US-08-452-843-5  
Description: (1-9) from US08452843.pep  
Perfect Score: 70  
Sequence: 1 FAMPNFQTL 9

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-gene:seq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 16.283; Variance 47.382; scale 0.344

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No.                          |
|------------|-------|-------------|--------|-------|-------------|------------------------------------|
| 1          | 70    | 100.0       | 9      | 18    | R89366      | Cw3 consensus peptide 2.10e-01     |
| 2          | 57    | 81.4        | 9      | 18    | R89367      | Cw3 consensus peptide 7.46e+00     |
| 3          | 48    | 68.6        | 845    | 13    | R70065      | Hepatitis B virus pol 7.85e+01     |
| 4          | 47    | 67.1        | 233    | 8     | R38869      | Sequence of type B fu 1.01e+02     |
| 5          | 47    | 67.1        | 233    | 11    | R60196      | Immunogenic fragment 1.01e+02      |
| 6          | 47    | 67.1        | 304    | 11    | R60197      | Immunogenic fragment 1.01e+02      |
| 7          | 47    | 67.1        | 304    | 11    | R60207      | Immunogenic fragment 1.01e+02      |
| 8          | 47    | 67.1        | 585    | 23    | W01671      | Influenza B/Harbin/77 1.01e+02     |
| 9          | 47    | 67.1        | 585    | 23    | W01672      | Influenza B/Harbin/77 1.01e+02     |
| 10         | 47    | 67.1        | 589    | 23    | W01673      | Influenza A/Shanghai/4 1.01e+02    |
| 11         | 47    | 67.1        | 592    | 23    | W01674      | Influenza A/Shanghai/4 1.01e+02    |
| 12         | 45    | 64.3        | 531    | 19    | R87615      | Rat N-acetylglucosamin 1.67e+02    |
| 13         | 45    | 64.3        | 531    | 19    | R87616      | Rat N-acetylglucosamin 1.67e+02    |
| 14         | 45    | 64.3        | 531    | 25    | R48994      | Human glycosyltransferase 1.67e+02 |
| 15         | 45    | 64.3        | 536    | 19    | R97614      | Human N-acetylglucosamin 1.67e+02  |
| 16         | 45    | 64.3        | 536    | 25    | R97615      | Human N-acetylglucosamin 1.67e+02  |
| 17         | 45    | 64.3        | 727    | 1     | W24014      | Rat N-acetylglucosamin 1.67e+02    |
| 18         | 45    | 64.3        | 4655   | 29    | W43312      | Human placental calci 1.67e+02     |

|    |    |      |      |    |        |                       |          |
|----|----|------|------|----|--------|-----------------------|----------|
| 19 | 45 | 64.3 | 4655 | 29 | W43313 | Human kidney calcium  | 1.67e+02 |
| 20 | 45 | 64.3 | 4655 | 29 | W43314 | Human calcium sensor  | 1.67e+02 |
| 21 | 45 | 64.3 | 4655 | 29 | W43315 | Human parathyroid cal | 1.67e+02 |
| 22 | 45 | 64.3 | 4655 | 17 | R97208 | Human calcium sensor  | 1.67e+02 |
| 23 | 45 | 64.3 | 4655 | 17 | R97209 | Human placental calci | 1.67e+02 |
| 24 | 45 | 64.3 | 4655 | 17 | R97210 | Human kidney calcium  | 1.67e+02 |
| 25 | 45 | 64.3 | 4655 | 17 | R97211 | Human parathyroid cal | 1.67e+02 |
| 26 | 44 | 62.9 | 277  | 7  | R37312 | Non-glycosylated REPI | 2.14e+02 |
| 27 | 44 | 62.9 | 955  | 28 | W31363 | Cell membrane proton- | 2.14e+02 |
| 28 | 44 | 62.9 | 1138 | 2  | R06461 | BTG51245 protoxin.    | 2.14e+02 |
| 29 | 43 | 61.4 | 326  | 4  | R21831 | Sequence encoded by b | 2.74e+02 |
| 30 | 43 | 61.4 | 385  | 29 | W55311 | H. pylori ORF 06p103  | 2.74e+02 |
| 31 | 43 | 61.4 | 428  | 25 | W26603 | Guar phosphomannose   | 2.74e+02 |
| 32 | 43 | 61.4 | 448  | 19 | R99425 | Brugia pahangi beta t | 2.74e+02 |
| 33 | 43 | 61.4 | 448  | 19 | R99423 | Diofilaria immitis b  | 2.74e+02 |
| 34 | 43 | 61.4 | 448  | 19 | R99424 | Onchocerca volvulus b | 2.74e+02 |
| 35 | 43 | 61.4 | 451  | 18 | R99739 | Retinoid X receptor i | 2.74e+02 |
| 36 | 43 | 61.4 | 469  | 20 | W03448 | Farnesoid-activated r | 2.74e+02 |
| 37 | 43 | 61.4 | 472  | 29 | W40072 | Human retinoid recept | 2.74e+02 |
| 38 | 43 | 61.4 | 484  | 18 | R99735 | Retinoid X receptor i | 2.74e+02 |
| 39 | 43 | 61.4 | 931  | 19 | W04867 | Transferrin binding p | 2.74e+02 |
| 40 | 43 | 61.4 | 933  | 7  | R33212 | Sequence of the p100  | 2.74e+02 |
| 41 | 43 | 61.4 | 1513 | 8  | R43253 | p190 protein.         | 2.74e+02 |
| 42 | 43 | 61.4 | 2861 | 27 | W27227 | Human TRIO phosphopro | 2.74e+02 |
| 43 | 42 | 60.0 | 542  | 27 | W32305 | Arabidopsis thaliana  | 3.50e+02 |
| 44 | 42 | 60.0 | 1162 | 23 | W19115 | Murine long form Ob r | 3.50e+02 |
| 45 | 42 | 60.0 | 1220 | 26 | W34500 | Obesity receptor D pr | 3.50e+02 |

## ALIGNMENTS

RESULT 1  
ID R89366 standard; peptide: 9 AA.

AC R89366;  
DE 18-SEP-1996 (first entry)  
DT Cw3 consensus peptide derived immunogenic peptide #1.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN W0603140-A1.  
PD 08-FEB-1996.  
PE 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-NOV-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI; 96-116784/12.  
PT Compan. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2: Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g. the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 100.0%; Score 70; DB 18; Length 9;  
Best Local Similarity 100.0%; Pred. No. 2.10e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 fampnfctl 9  
QY 1 FAMPNFQTL 9

RESULT 2  
ID R89367 standard; peptide: 9 AA.  
AC R89367;  
DE 18-SEP-1996 (first entry)  
DE Cw3 consensus peptide derived immunogenic peptide #2.  
KW immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; 009234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI: 96-116784/12.  
PT Compos. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications  
PT Claim 2; Page 26; 32pp; English.  
PS The sequences given in R89362-82 are immunogenic peptides which were used in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which allows binding of more than one HLA molecule. It pref. comprises two conserved residues, a first at the 2nd position from the N-terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also useful in compositions for in vivo and ex vivo therapeutic and diagnostic applications, e.g. the treatment of cancer and viral infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA: hepatitis B and C.

Query Match: 81.4%; Score 57; DB 18; Length 9;  
Best Local Similarity 88.9%; Pred. No. 7.46e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 famprf1 9  
1 FAMPRFOTL 9

QY 1 FAMPRFOTL 9

RESULT 3  
ID R70065 standard; protein: 845 AA.  
AC R70065;  
DE 06-OCT-1995 (first entry)  
DE Hepatitis B virus polymerase protein.  
KM Hepatitis B virus polymerase; cytotoxic T cell response; prophylactic; vaccine; chronic; acute HBV infection; carrier.  
OS Hepatitis B virus.  
PN WO9503777-A.  
PD 09-FEB-1995.  
PF 01-AUG-1994; 008685.  
PR 02-AUG-1993; US-100870.  
PA (SCRI) SCRIPPS RES INST.  
PI Chisari FV;  
DR WPI: 95-082004/11.  
PT New peptides inducing cytotoxic T lymphocytes to hepatitis B virus - are regions of HB polymerase protein, for treating acute and chronic infections  
PS Disclosure: Page 50-52; 85pp; English.  
CC The amino acid sequence of the Hepatitis B virus (HBV) polymerase (HBpol) protein. The sequence was used to generate a series of peptides (R70044-59) which induce cytotoxic T cell (CTL) responses against cells infected with HBV. The HBpol peptides can be used, prophylactically as vaccines, together with or conjugated to, epitopes from other HBV sequences that elicit T cell responses to HBV (see R70060-64). The peptides can be used, particularly ex vivo, to stimulate CTL cells. These cells can be reintroduced into patients who have chronic or acute HBV infections or are carriers, especially in treatments to prevent conversion from acute to chronic infections.

SQ Sequence 845 AA:  
Query Match 68.6%; Score 48; DB 13; Length 845;  
Best Local Similarity 66.7%; Pred. No. 7.85e+01;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 409 famprf1 417  
1 FAMPRFOTL 9

QY 1 FAMPRFOTL 9

RESULT 4  
ID R38869 standard; protein: 233 AA.  
AC R38869;  
DE 04-FEB-1994 (first entry)  
DE Sequence of type B fusion protein NSI(1-42)HA2(41-223)  
KW vaccine; influenza virus; haemagglutinin subunit; HA2.  
OS Synthetic.  
PN WO9315763-A.  
PD 19-AUG-1993.  
PF 18-FEB-1993; 001451.  
PR 18-FEB-1992; US-837773.  
PA (SMIK) SMITHKLINE BEECHAM CORP.  
PI Dillon SB, Scott M, Shatzman A;  
DR WPI: 93-272565/34.  
PT Vaccine against influenza A and B - contg. haemagglutinin 2 sub-unit of virus, and conferring multi-strain immunity  
PS Claim 14; Page 58; 99pp; English.  
CC Proteins of the invention are derived from the HA2 subunit of a haemagglutinin (HA) protein, e.g., from a H3N2 subtype virus. Among H3N2 subtype strains of influenza A include A/Udorn and A/Victoria viruses. Examples are Aas 1-221 and 77-221 of a selected H3HA2 subunit. Fusion proteins are also claimed, which include a protein derived from a H3N2 subtype virus fused in frame with, e.g., the CC NSI portion derived from a H1N1 subtype virus, A/PR/8/34 (04/7360). The NSI portion may comprise residues 1-42 or 1-81 of H1NSI. Alternatively, the HA2 fragment may be fused to a portion of the CC NSI peptide derived from a selected type A virus, e.g., an H3 subtype virus (H3HA2), or a type B (BHA2) virus. The preferred type CC B influenza virus is human virus strain B/Lee/40. A type B fusion protein is NSI(1-42)HA2(41-223).  
SQ Sequence 233 AA:

Query Match 67.1%; Score 47; DB 8; Length 233;  
Best Local Similarity 44.4%; Pred. No. 1.01e+02;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 172 famprf1 180  
1 FAMPRFOTL 9

QY 1 FAMPRFOTL 9

RESULT 5  
ID R60196 standard; protein: 233 AA.  
AC R60196;  
DE 28-MAR-1995 (first entry)  
DE Immunogenic fragment of influenza haemagglutinin (fusion protein).  
KM Antigen; immunogen; vaccine; influenza; fusion protein; immunity;  
KW haemagglutinin; neuraminidase; flu.  
OS Influenza virus.  
PN WO9417826-A.  
PD 18-AUG-1994.  
PF 01-FEB-1994; 001149.  
PR 01-FEB-1993; US-013415.  
PR 18-AUG-1993; US-108914.  
PR 05-NOV-1993; US-149150.  
PA (SMIK) SMITHKLINE BEECHAM CORP.  
PI Dillon S, Kane J, Scott M, Shatzman A;  
DR WPI: 94-279392/34.  
DR N-PSDB: 070192.  
PT Vaccines against multi strain influenza virus infection - protect against influenza A and B

PS Claim 8: Page 70-71: 151pp; English.  
 CC A vaccine comprising an immunogenic fragment of the HA2 subunit of  
 CC the influenza haemagglutinin (HA) protein from type A subtype IV and  
 CC type B IV may be used for stimulating protection in animals against  
 CC infection with influenza virus. The vaccine confers multi-strain  
 CC immunity against strains IV A and IV B. The vaccines may be  
 CC recombinantly produced, optionally as fusion proteins. In this  
 CC sequence the N-terminal 42 amino acids are derived from the influenza  
 CC NS1 protein and the remainder of the sequence comprises amino acids  
 CC 41-223 of the HA2 subunit of the BUHA2 subtype of influenza.  
 SQ Sequence 233 AA;

Query Match  
 Best Local Similarity 44.4%; Score 47; DB 11; Length 233;  
 Pred. No. 1.01e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 172 fs|ptfda| 180  
 1:1 1:1  
 1 FAMPNFOTL 9

RESULT 6  
 ID R60197 standard; Protein: 304 AA.  
 AC R60197;  
 DT 28-MAR-1995 (first entry)  
 DE Immunogenic fragment of influenza haemagglutinin (fusion protein).  
 KW Antigen; Immunogen; Vaccine; Influenza; Fusion protein; Immunity;  
 KM haemagglutinin; neuraminidase; flu.  
 OS Influenza virus.  
 PN WO9417826-A.  
 PD 18-AUG-1994.  
 PF 01-FEB-1994: U01149.  
 PR 01-FEB-1993: US-013415.  
 PR 18-AUG-1993: US-108914.  
 PR 05-NOV-1993: US-149150.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Dillon S, Kane J, Scott M, Shatzman A;  
 DR WPI: 94-279392/34.  
 DR N-PSDB: 070193.  
 PT Vaccines against multi strain influenza virus infection - protect  
 PT against influenza A and B  
 PS Claim 9: Page 111-112: 151pp; English.  
 CC A vaccine comprising an immunogenic fragment of the HA2 subunit of  
 CC the influenza haemagglutinin (HA) protein from type A subtype IV and  
 CC type B IV may be used for stimulating protection in animals against  
 CC infection with influenza virus. The vaccine confers multi-strain  
 CC immunity against strains IV A and IV B. The vaccines may be  
 CC recombinantly produced, optionally as fusion proteins. In this  
 CC sequence the N-terminal 81 amino acids are derived from the influenza  
 CC NS1 protein and the remainder of the sequence comprises amino acids  
 CC 1-223 of the HA2 subunit of the BUHA2 subtype of influenza.  
 SQ Sequence 304 AA;

Query Match  
 Best Local Similarity 44.4%; Score 47; DB 11; Length 304;  
 Pred. No. 1.01e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 243 fs|ptfda| 251  
 1:1 1:1  
 1 FAMPNFOTL 9

RESULT 7  
 ID R60207 standard; Protein: 304 AA.  
 AC R60207;  
 DT 28-MAR-1995 (first entry)  
 DE Immunogenic fragment of influenza haemagglutinin (fusion protein).  
 KW Antigen; Immunogen; Vaccine; Influenza; Fusion protein; Immunity;  
 KM haemagglutinin; neuraminidase; flu.  
 OS Influenza virus.  
 PN WO9417826-A.  
 PD 18-AUG-1994.  
 PF 01-FEB-1994: U01149.

PR 01-FEB-1993: US-013415.  
 PR 18-AUG-1993: US-108914.  
 PR 05-NOV-1993: US-149150.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Dillon S, Kane J, Scott M, Shatzman A;  
 DR WPI: 94-279392/34.  
 DR N-PSDB: 070194.  
 PT Vaccines against multi strain influenza virus infection - protect  
 PT against influenza A and B  
 PS Claim 10: Page 107-108: 151pp; English.  
 CC A vaccine comprising an immunogenic fragment of the HA2 subunit of  
 CC the influenza haemagglutinin (HA) protein from type A subtype IV and  
 CC type B IV may be used for stimulating protection in animals against  
 CC infection with influenza virus. The vaccine confers multi-strain  
 CC immunity against strains IV A and IV B. The vaccines may be  
 CC recombinantly produced, optionally as fusion proteins. In this  
 CC sequence the N-terminal 81 amino acids are derived from the influenza  
 CC NS1 protein and the remainder of the sequence comprises amino acids  
 CC 1-223 of the HA2 subunit of the BUHA2 (met-leu) subtype of influenza.  
 SQ Sequence 304 AA;

Query Match  
 Best Local Similarity 44.4%; Score 47; DB 11; Length 304;  
 Pred. No. 1.01e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 243 fs|ptfda| 251  
 1:1 1:1  
 1 FAMPNFOTL 9

RESULT 8  
 ID W01671 standard; Protein: 585 AA.  
 AC W01671;  
 DT 19-AUG-1997 (first entry)  
 DE Influenza B/Panama/45/90 recombinant haemagglutinin protein.  
 KW primer; PCR; polymerase chain reaction; universal; amplify; HA;  
 KM haemagglutinin; recombinant production; baculovirus expression system;  
 KW vaccine; insect cell culture.  
 OS Synthetic.  
 FH Key  
 FT peptide  
 FT Location/Qualifiers  
 FT 1..17  
 FT /label= AcNPV\_61K\_protein\_signal\_sequence  
 FT 18..568  
 FT /label= mature\_recombinant\_haemagglutinin  
 FT protein  
 FT W09637624-A1.  
 PN 28-NOV-1996.  
 PD 26-MAY-1995: U06750.  
 PR 26-MAY-1995: WO-006750.  
 PA (MICR-) MICROGENESIS INC.  
 PA (MGPM-) MG-PMC LLC.  
 PI Hackett CS, Smith GE, Volovitz F, Voznesensky AI;  
 PI Wilkinson BE;  
 DR WPI: 97-021228/02.  
 DR N-PSDB: T59214.  
 PT Recombinant influenza haemagglutinin produced in baculovirus system  
 PT - avoids problems of growing virus in eggs and produces stable,  
 PT un-cleaved protein useful in vaccines  
 PS Example 3: Page 75-77: 107pp; English.  
 CC Recombinant influenza haemagglutinin (HA) expressed in a  
 CC baculovirus expression system in cultured insect cells, allows vaccine  
 CC production without the need to grow virus in eggs. A purer, less  
 CC allergenic product is obtained and antigen drift caused by passages  
 CC through eggs is avoided. There is no need for viral inactivation or  
 CC organic solvent extr. of viral membrane components and vaccines can be  
 CC prepd. rapidly and cost effectively from primary sources of infection.  
 CC Recombinant HA is more stable (esp. for B strains) than HA1/HA2 complexes  
 CC and maintain correct folding during purification and storage. The present  
 CC sequence shows the N-terminal end of the HA protein for influenza  
 CC B/Panama/45/90 (sequence range 1-434).  
 SQ Sequence 585 AA;

Query Match  
 Best Local Similarity 44.4%; Score 47; DB 23; Length 585;  
 Pred. No. 1.01e+02;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 524 fs1ptfda1.532

OY 1 FAMNFOFL 9

# RESULT 9

ID W01675 standard; Protein: 586 AA.

AC W01675; 19-AUG-1997 (first entry)

DE Influenza A/B/Hardin/7/94 recombinant haemagglutinin protein.

KM primer: PCR; polymerase chain reaction; universal; amplify; HA;

KW haemagglutinin; recombinant production; baculovirus expression system;

OS Synthetic; insect cell culture; ss.

FT Key Location/Qualifiers

FT Peptide 1..17 /label- ACNPV\_61K\_protein\_signal\_peptide

FT protein 18..569

FTM protein /label- mature\_rHA

PN W0637624-A1.

PD 28-NOV-1996.

PE 26-MAY-1995; U06750.

PR 26-MAY-1995; WO-U06750.

PA (MICR-) MICROGENESYS INC.

PI (MGPM-) MG-PMC LLC.

PI Hackett CS, Smith GE, Volvovitz F, Voznesensky AI;

PI Wilkinson BE;

DR WPI: 97-021228/02.

DR N-PSDB: T59218.

PT Recombinant influenza haemagglutinin produced in baculovirus system

PT - avoids problems of growing virus in eggs and produces stable,

PT un-cleaved protein useful in vaccines

PS Example 12; Page 87-88; 107pp; English.

CC Recombinant influenza haemagglutinin (HA) expressed in a

CC baculovirus expression system in cultured insect cells, allows vaccine

CC production without the need to grow virus in eggs. A purer, less

CC allergenic product is obtained and antigen drift caused by passages

CC through eggs is avoided. There is no need for viral inactivation or

CC organic solvent extn. of viral membrane components and vaccines can be

CC prep. rapidly and cost effectively from primary sources of infection.

CC Recombinant HA is more stable (esp. for B strains) than HA1/HA2 complexes

CC and maintain correct folding during purification and storage. The present

CC sequence shows the N-terminal end of the HA gene for influenza

CC B/Hardin/7/94 which is used in a trivalent influenza virus vaccine.

CC Sequence 586 AA;

SO

Query Match

Best Local Similarity 44.4%; Score 47; DB 23; Length 586;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 525 fs1ptfda1.533

OY 1 FAMNFOFL 9

# RESULT 10

ID W01672 standard; Protein: 589 AA.

AC W01672; 19-AUG-1997 (first entry)

DE Influenza A/B/Netherlands/13/94 recombinant haemagglutinin protein.

KM primer: PCR; polymerase chain reaction; universal; amplify; HA;

KW haemagglutinin; recombinant production; baculovirus expression system;

OS Synthetic; insect cell culture.

FT Key Location/Qualifiers

FT Peptide 1..18 /label- ACNPV\_61K\_protein\_signal\_sequence

FT protein 19..571

FTM protein /label- mature\_rHA

PN W0637624-A1.

PD 28-NOV-1996.

PE 26-MAY-1995; U06750.

PR 26-MAY-1995; WO-U06750.

PA (MICR-) MICROGENESYS INC.

PI (MGPM-) MG-PMC LLC.

PI Hackett CS, Smith GE, Volvovitz F, Voznesensky AI;

PI Wilkinson BE;

DR WPI: 97-021228/02.

DR N-PSDB: T59217.

PT Recombinant influenza haemagglutinin produced in baculovirus system

PT - avoids problems of growing virus in eggs and produces stable,

PT un-cleaved protein useful in vaccines

PS Disclosure: Page 84-85; 107pp; English.

CC Recombinant influenza haemagglutinin (HA) expressed in a

CC baculovirus expression system in cultured insect cells, allows vaccine

CC production without the need to grow virus in eggs. A purer, less

CC allergenic product is obtained and antigen drift caused by passages

CC through eggs is avoided. There is no need for viral inactivation or

CC organic solvent extn. of viral membrane components and vaccines can be

CC prep. rapidly and cost effectively from primary sources of infection.

CC Recombinant HA is more stable (esp. for B strains) than HA1/HA2 complexes

CC and maintain correct folding during purification and storage. The present

CC sequence shows the N-terminal end of the HA gene for influenza

CC A/Shanghai/4/94.

PF 26-MAY-1995; U06750.

PR 26-MAY-1995; WO-U06750.

PA (MICR-) MICROGENESYS INC.

PI (MGPM-) MG-PMC LLC.

PI Hackett CS, Smith GE, Volvovitz F, Voznesensky AI;

PI Wilkinson BE;

DR WPI: 97-021228/02.

DR N-PSDB: T59215.

PT Recombinant influenza haemagglutinin produced in baculovirus system

PT - avoids problems of growing virus in eggs and produces stable,

PT un-cleaved protein useful in vaccines

PS Disclosure: Page 78-79; 107pp; English.

CC Recombinant influenza haemagglutinin (HA) expressed in a

CC baculovirus expression system in cultured insect cells, allows vaccine

CC production without the need to grow virus in eggs. A purer, less

CC allergenic product is obtained and antigen drift caused by passages

CC through eggs is avoided. There is no need for viral inactivation or

CC organic solvent extn. of viral membrane components and vaccines can be

CC prep. rapidly and cost effectively from primary sources of infection.

CC Recombinant HA is more stable (esp. for B strains) than HA1/HA2 complexes

CC and maintain correct folding during purification and storage. The present

CC sequence shows the N-terminal end of the HA gene for influenza

CC B/Netherlands/13/94.

SO Sequence 589 AA;

Query Match

Best Local Similarity 44.4%; Score 47; DB 23; Length 589;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 528 fs1ptfda1.536

OY 1 FAMNFOFL 9

# RESULT 11

ID W01674 standard; Protein: 592 AA.

AC W01674; 19-AUG-1997 (first entry)

DE Influenza A/Shanghai/4/94 recombinant haemagglutinin protein.

KM primer: PCR; polymerase chain reaction; universal; amplify; HA;

KW haemagglutinin; recombinant production; baculovirus expression system;

OS Synthetic; insect cell culture.

FT Key Location/Qualifiers

FT Peptide 1..18 /label- ACNPV\_61K\_protein\_signal\_peptide

FT protein 19..574

FTM protein /label- mature\_rHA

PN W0637624-A1.

PD 28-NOV-1996.

PE 26-MAY-1995; U06750.

PR 26-MAY-1995; WO-U06750.

PA (MICR-) MICROGENESYS INC.

PI (MGPM-) MG-PMC LLC.

PI Hackett CS, Smith GE, Volvovitz F, Voznesensky AI;

PI Wilkinson BE;

DR WPI: 97-021228/02.

DR N-PSDB: T59217.

PT Recombinant influenza haemagglutinin produced in baculovirus system

PT - avoids problems of growing virus in eggs and produces stable,

PT un-cleaved protein useful in vaccines

PS Disclosure: Page 84-85; 107pp; English.

CC Recombinant influenza haemagglutinin (HA) expressed in a

CC baculovirus expression system in cultured insect cells, allows vaccine

CC production without the need to grow virus in eggs. A purer, less

CC allergenic product is obtained and antigen drift caused by passages

CC through eggs is avoided. There is no need for viral inactivation or

CC organic solvent extn. of viral membrane components and vaccines can be

CC prep. rapidly and cost effectively from primary sources of infection.

CC Recombinant HA is more stable (esp. for B strains) than HA1/HA2 complexes

CC and maintain correct folding during purification and storage. The present

CC sequence shows the N-terminal end of the HA gene for influenza

CC A/Shanghai/4/94.

SQ Sequence 592 AA:

Query Match 67.1%; Score 47; DB 23; Length 592;

Best Local Similarity 44.4%; Pred. No. 1.01e+02; Mismatches 1; Indels 0; Gaps 0;

DB 531 fsiptfasi 539

OY 1 FAMPMFQTL 9

ID R97615 standard; Protein; 531 AA.

AC R97615;

DE 21-NOV-1996 (first entry)

KW Rat N-acetylglucosaminyl transferase-III, inhibitor of metastasis.

KW N-acetylglucosaminyl transferase-III; metastasis inhibitor; cancer;

OS Rattus sp.

PN J08109139-A.

PD 30-APR-1996.

PF 12-OCT-1994; 271802.

PR 12-OCT-1994; JP-271802.

PA (TAKI) TAKARA SHUZO CO LTD.

DR MPI; 96-263788/27.

PT Inhibitor of cancer metastasis - contains N-acetylglucosaminyl

PS transferase-III isolated from rat kidneys

CC Claim 3; Page 8-10; 12pp; Japanese.

CC R97614 and R97615 are two clones of rat N-acetylglucosaminyl

CC transferase-III (Gnt-III). These clones are useful to reinforce

CC the activity of Gnt-III present in cancer cells and cells surrounding

CC cancer cells for the inhibition of metastasis, preventing the spread

CC of cancer cells to tissues other than that of their origin. The

CC clones were isolated from a rat kidney homogenate.

SQ Sequence 531 AA:

Query Match 64.3%; Score 45; DB 19; Length 531;

Best Local Similarity 57.1%; Pred. No. 1.67e+02; Mismatches 0; Indels 0; Gaps 0;

DB 385 ytmprf 391

OY 1 FAMPMFQ 7

ID R48994 standard; Protein; 531 AA.

AC R48994;

DE 04-SEP-1994 (first entry)

KW Human glycosyltransferase Gnt-III protein.

KW Glycosyltransferase; enzyme; cancer diagnosis; ss.

OS Homo sapiens.

PN EP-585083-A.

PD 02-MAR-1994.

PF 20-AUG-1993; 306628.

PR 21-AUG-1992; JP-243984.

PA (TAKI) TAKARA SHUZO CO LTD.

DR MPI; 94-067563/09.

PT New gene for human glycosyltransferase Gnt-III - and related

PS vectors and transformed cells, useful in diagnosis of cancer

CC Claim 3; Page 10; 14pp; English.

CC This glycosyltransferase protein is human UDP-N-acetylglucosamine;

CC beta-D-mannoside-beta1-4N-acetylglucosaminyltransferase. It may

CC be expressed recombinantly in host cells and used in cancer diagnosis.

SQ Sequence 531 AA:

OY 1 FAMPMFQ 7

ID W24015 standard; Protein; 531 AA.

AC W24015;

DE 10-FEB-1998 (first entry)

KW Human N-acetylglucosaminyl transferase III.

KW N-acetylglucosaminyl transferase III; Gnt-III; rat; human;

KW virus; replication; inhibitor; hepatitis B; hepatitis C; HIV;

KW viral disease; human immunodeficiency virus.

OS Homo sapiens.

PN MO9718836-A1.

PD 29-MAR-1997.

PF 17-JUL-1996; J01986.

PR 17-NOV-1995; JP-322474.

PA (TAKI) TAKARA SHUZO CO LTD.

DR MPI; 97-297877/27.

PT Virus replication inhibitor containing N-acetylglucosaminyl

PS transferase III or its gene - useful in treatment of hepatitis B, C

CC Claim 6; Page 22-26; 12pp; Japanese.

CC The present sequence represents human N-acetylglucosaminyl transferase

CC III (Gnt-III). The present specification describes a virus replication

CC inhibitor which contains the gene encoding Gnt-III as an active

CC ingredient. The virus replication inhibitor is used against hepatitis B,

CC C and HIV. It is used for the treatment of viral diseases.

SQ Sequence 531 AA:

Query Match 64.3%; Score 45; DB 25; Length 531;

Best Local Similarity 57.1%; Pred. No. 1.67e+02; Mismatches 0; Indels 0; Gaps 0;

DB 385 ytmprf 391

OY 1 FAMPMFQ 7

ID R97614 standard; Protein; 536 AA.

AC R97614;

DE 21-NOV-1996 (first entry)

KW Rat N-acetylglucosaminyl transferase-III, inhibitor of metastasis.

KW N-acetylglucosaminyl transferase-III; metastasis inhibitor; cancer;

OS Rattus sp.

PN J08109139-A.

PD 30-APR-1996.

PF 12-OCT-1994; 271802.

PR 12-OCT-1994; JP-271802.

PA (TAKI) TAKARA SHUZO CO LTD.

DR MPI; 96-263788/27.

PT Inhibitor of cancer metastasis - contains N-acetylglucosaminyl

PS transferase-III isolated from rat kidneys

CC Claim 3; Page 6-8; 12pp; Japanese.

CC R97614 and R97615 are two clones of rat N-acetylglucosaminyl

CC transferase-III (Gnt-III). These clones are useful to reinforce

CC the activity of Gnt-III present in cancer cells and cells surrounding

CC cancer cells for the inhibition of metastasis, preventing the spread

CC of cancer cells to tissues other than that of their origin. The

CC clones were isolated from a rat kidney homogenate.

SQ Sequence 536 AA:

OY 1 FAMPMFQ 7

ID W24015 standard; Protein; 531 AA.

AC W24015;

DE 10-FEB-1998 (first entry)

KW Human N-acetylglucosaminyl transferase III.

KW N-acetylglucosaminyl transferase III; Gnt-III; rat; human;

KW virus; replication; inhibitor; hepatitis B; hepatitis C; HIV;

KW viral disease; human immunodeficiency virus.

OS Homo sapiens.

PN MO9718836-A1.

PD 29-MAR-1997.

PF 17-JUL-1996; J01986.

PR 17-NOV-1995; JP-322474.

PA (TAKI) TAKARA SHUZO CO LTD.

DR MPI; 97-297877/27.

PT Virus replication inhibitor containing N-acetylglucosaminyl

PS transferase III or its gene - useful in treatment of hepatitis B, C

CC Claim 6; Page 22-26; 12pp; Japanese.

CC The present sequence represents human N-acetylglucosaminyl transferase

CC III (Gnt-III). The present specification describes a virus replication

CC inhibitor which contains the gene encoding Gnt-III as an active

CC ingredient. The virus replication inhibitor is used against hepatitis B,

CC C and HIV. It is used for the treatment of viral diseases.

SQ Sequence 531 AA:

Query Match 64.3%; Score 45; DB 25; Length 531;

Best Local Similarity 57.1%; Pred. No. 1.67e+02; Mismatches 0; Indels 0; Gaps 0;

DB 385 ytmprf 391

OY 1 FAMPMFQ 7

ID R97614 standard; Protein; 536 AA.

AC R97614;

DE 21-NOV-1996 (first entry)

KW Rat N-acetylglucosaminyl transferase-III, inhibitor of metastasis.

KW N-acetylglucosaminyl transferase-III; metastasis inhibitor; cancer;

OS Rattus sp.

PN J08109139-A.

PD 30-APR-1996.

PF 12-OCT-1994; 271802.

PR 12-OCT-1994; JP-271802.

PA (TAKI) TAKARA SHUZO CO LTD.

DR MPI; 96-263788/27.

PT Inhibitor of cancer metastasis - contains N-acetylglucosaminyl

PS transferase-III isolated from rat kidneys

CC Claim 3; Page 6-8; 12pp; Japanese.

CC R97614 and R97615 are two clones of rat N-acetylglucosaminyl

CC transferase-III (Gnt-III). These clones are useful to reinforce

CC the activity of Gnt-III present in cancer cells and cells surrounding

CC cancer cells for the inhibition of metastasis, preventing the spread

CC of cancer cells to tissues other than that of their origin. The

CC clones were isolated from a rat kidney homogenate.

SQ Sequence 536 AA:

OY 1 FAMPMFQ 7

ID W24015 standard; Protein; 531 AA.

AC W24015;

DE 10-FEB-1998 (first entry)

KW Human N-acetylglucosaminyl transferase III.

KW N-acetylglucosaminyl transferase III; Gnt-III; rat; human;

KW virus; replication; inhibitor; hepatitis B; hepatitis C; HIV;

KW viral disease; human immunodeficiency virus.

OS Homo sapiens.

PN MO9718836-A1.

PD 29-MAR-1997.

PF 17-JUL-1996; J01986.

PR 17-NOV-1995; JP-322474.

PA (TAKI) TAKARA SHUZO CO LTD.

DR MPI; 97-297877/27.

PT Virus replication inhibitor containing N-acetylglucosaminyl

PS transferase III or its gene - useful in treatment of hepatitis B, C

CC Claim 6; Page 22-26; 12pp; Japanese.

CC The present sequence represents human N-acetylglucosaminyl transferase

CC III (Gnt-III). The present specification describes a virus replication

CC inhibitor which contains the gene encoding Gnt-III as an active

CC ingredient. The virus replication inhibitor is used against hepatitis B,

CC C and HIV. It is used for the treatment of viral diseases.

SQ Sequence 531 AA:

Query Match 64.3%; Score 45; DB 25; Length 531;

Best Local Similarity 57.1%; Pred. No. 1.67e+02; Mismatches 0; Indels 0; Gaps 0;

DB 385 ytmprf 391

OY 1 FAMPMFQ 7

ID R97614 standard; Protein; 536 AA.

AC R97614;

DE 21-NOV-1996 (first entry)

KW Rat N-acetylglucosaminyl transferase-III, inhibitor of metastasis.

KW N-acetylglucosaminyl transferase-III; metastasis inhibitor; cancer;

OS Rattus sp.

PN J08109139-A.

PD 30-APR-1996.

PF 12-OCT-1994; 271802.

PR 12-OCT-1994; JP-271802.

PA (TAKI) TAKARA SHUZO CO LTD.

DR MPI; 96-263788/27.

PT Inhibitor of cancer metastasis - contains N-acetylglucosaminyl

PS transferase-III isolated from rat kidneys

CC Claim 3; Page 6-8; 12pp; Japanese.

CC R97614 and R97615 are two clones of rat N-acetylglucosaminyl

CC transferase-III (Gnt-III). These clones are useful to reinforce

CC the activity of Gnt-III present in cancer cells and cells surrounding

CC cancer cells for the inhibition of metastasis, preventing the spread

CC of cancer cells to tissues other than that of their origin. The

CC clones were isolated from a rat kidney homogenate.

SQ Sequence 536 AA:

Sun Sep 13 10:56:54 1998

US-08-452-843-5.rag

Page 6

OY 1 FAMNFO 7

Search completed: Fri Sep 11 12:53:49 1998  
Job time : 16.secs.

---

\*\*\*\*\*  
 M P E R E H  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSPrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Sep 11 12:54:06 1998; Maspar time 3.34 Seconds  
 Tabular output not generated. 98.507 Million cell updates/sec

Title: >US-08-452-843-5  
 Description: (1-9) from US08452843.pep  
 Perfect Score: 70  
 Sequence: 1 FAMPNFOTL 9

Scoring table: PAM 150  
 Gap 15

Searched: 120441 segs, 36531193 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: pir56  
 1:pir1 2:pir2 3:pir3 4:pir4 5:nrl3d

Statistics: Mean 23.027; Variance 31.559; scale 0.730

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Match | Length | DB | ID      | Description           | Pred. No. |
|------------|-------|-------|--------|----|---------|-----------------------|-----------|
| 1          | 53    | 75.7  | 744    | 2  | S65669  | biotin sulfoxide redu | 3.21e+00  |
| 2          | 52    | 74.3  | 87     | 2  | S72830  | hypothetical protein  | 4.99e+00  |
| 3          | 51    | 72.9  | 217    | 2  | S64359  | ribosomal protein S5  | 7.70e+00  |
| 4          | 51    | 72.9  | 556    | 1  | JDVIL64 | DNA-directed DNA poly | 7.70e+00  |
| 5          | 51    | 72.9  | 879    | 1  | JDVLC   | DNA-directed DNA poly | 7.70e+00  |
| 6          | 51    | 72.9  | 881    | 1  | JDVLS   | DNA-directed DNA poly | 7.70e+00  |
| 7          | 51    | 72.9  | 883    | 1  | JDVLC2  | DNA-directed DNA poly | 7.70e+00  |
| 8          | 51    | 72.9  | 884    | 1  | JDVLS9  | DNA-directed DNA poly | 7.70e+00  |
| 9          | 51    | 72.9  | 884    | 1  | JDVLS8  | DNA-directed DNA poly | 7.70e+00  |
| 10         | 51    | 72.9  | 884    | 1  | JDVLT7  | DNA-directed DNA poly | 7.70e+00  |
| 11         | 49    | 70.0  | 310    | 2  | B47050  | glua 3'-region hypoch | 1.81e+01  |
| 12         | 49    | 70.0  | 767    | 2  | S68148  | AMP deaminase (EC 3.5 | 1.81e+01  |
| 13         | 49    | 70.0  | 774    | 2  | S68147  | AMP deaminase (EC 3.5 | 1.81e+01  |
| 14         | 49    | 70.0  | 776    | 2  | S68146  | AMP deaminase (EC 3.5 | 1.81e+01  |
| 15         | 48    | 68.6  | 205    | 2  | I49364  | protein tyrosine phos | 2.75e+01  |
| 16         | 48    | 68.6  | 309    | 2  | S40748  | hypothetical protein  | 2.75e+01  |
| 17         | 48    | 68.6  | 376    | 5  | IHC12   | hemocyanin chain a su | 2.75e+01  |
| 18         | 48    | 68.6  | 376    | 5  | IHC2    | hemocyanin chain a su | 2.75e+01  |
| 19         | 48    | 68.6  | 376    | 5  | IHC32   | hemocyanin chain a su | 2.75e+01  |
| 20         | 48    | 68.6  | 376    | 5  | IHC42   | hemocyanin chain a su | 2.75e+01  |
| 21         | 48    | 68.6  | 376    | 5  | IHC52   | hemocyanin chain a su | 2.75e+01  |
| 22         | 48    | 68.6  | 376    | 5  | IHC22   | hemocyanin chain a su | 2.75e+01  |
| 23         | 48    | 68.6  | 564    | 1  | HMITVF9 | hemagglutinin precurs | 2.75e+01  |

|    |    |      |     |   |         |                       |          |
|----|----|------|-----|---|---------|-----------------------|----------|
| 24 | 48 | 68.6 | 564 | 1 | HMITVF3 | hemagglutinin precurs | 2.75e+01 |
| 25 | 48 | 68.6 | 564 | 1 | HMITVF6 | hemagglutinin precurs | 2.75e+01 |
| 26 | 48 | 68.6 | 564 | 1 | HMITVF5 | hemagglutinin precurs | 2.75e+01 |
| 27 | 48 | 68.6 | 564 | 1 | HMITVF8 | hemagglutinin precurs | 2.75e+01 |
| 28 | 48 | 68.6 | 564 | 1 | A46339  | hemagglutinin precurs | 2.75e+01 |
| 29 | 48 | 68.6 | 565 | 5 | IHCY1   | hemocyanin chain a (d | 2.75e+01 |
| 30 | 48 | 68.6 | 657 | 1 | BHLOA   | hemocyanin chain b -  | 2.75e+01 |
| 31 | 48 | 68.6 | 657 | 1 | BHLOB   | hemocyanin chain b -  | 2.75e+01 |
| 32 | 48 | 68.6 | 832 | 1 | S20752  | DNA-directed DNA poly | 2.75e+01 |
| 33 | 48 | 68.6 | 832 | 1 | JDVIVA  | DNA-directed DNA poly | 2.75e+01 |
| 34 | 48 | 68.6 | 832 | 1 | S47406  | DNA-directed DNA poly | 2.75e+01 |
| 35 | 48 | 68.6 | 832 | 1 | JDVICP  | DNA-directed DNA poly | 2.75e+01 |
| 36 | 48 | 68.6 | 832 | 1 | JDVLYB  | DNA-directed DNA poly | 2.75e+01 |
| 37 | 48 | 68.6 | 832 | 2 | S67505  | DNA-directed DNA poly | 2.75e+01 |
| 38 | 48 | 68.6 | 832 | 2 | S20757  | DNA-directed DNA poly | 2.75e+01 |
| 39 | 48 | 68.6 | 843 | 1 | JQ0229  | DNA-directed DNA poly | 2.75e+01 |
| 40 | 48 | 68.6 | 843 | 1 | S43491  | DNA-directed DNA poly | 2.75e+01 |
| 41 | 48 | 68.6 | 843 | 1 | JDVLYR  | DNA-directed DNA poly | 2.75e+01 |
| 42 | 48 | 68.6 | 843 | 2 | S35527  | DNA-directed DNA poly | 2.75e+01 |
| 43 | 48 | 68.6 | 843 | 1 | JDVLI1  | DNA-directed DNA poly | 2.75e+01 |
| 44 | 48 | 68.6 | 845 | 1 | JDVLYD  | DNA-directed DNA poly | 2.75e+01 |
| 45 | 48 | 68.6 | 845 | 1 | JDVLYS  | DNA-directed DNA poly | 2.75e+01 |

## ALIGNMENTS

RESULT 1  
 ENTRY S65669 #type complete  
 TITLE biotin sulfoxide reductase - Rhodobacter sphaeroides  
 ORGANISM #formal\_name Rhodobacter sphaeroides  
 DATE 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 13-Feb-1998

ACCESSIONS  
 REFERENCE S65669  
 #authors Pollock, V.V.; Barber, M.J.  
 #journal Arch. Biochem. Biophys. (1995) 318:322-332  
 #title Molecular cloning and expression of biotin sulfoxide reductase from Rhodobacter sphaeroides forma sp. denitrificans.  
 #accession S65669  
 #molecule\_type DNA  
 #residues 1-744 #label POL  
 ##cross-references EMBL:U08189; NID:9953223; PID:9953224  
 ##experimental\_source strain forma sp. denitrificans  
 ##note the authors translated the inflation codon GTG for residue 1 as Val

## GENETICS

#start\_codon GTG  
 KEYWORDS molybdenum; oxidoreductase  
 SUMMARY #length 744 #molecular\_weight 80266 #checksum 9898

Query Match 75.7%; Score 53; DB 2; Length 744;  
 Best local Similarity 55.6%; Pred. No. 3.21e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 519 FEMDPETF 527  
 QY 1 FAMPNFOTL 9

RESULT 2  
 ENTRY S72830 #type complete  
 TITLE hypothetical protein B1620\_F1.14 - Mycobacterium leprae  
 ORGANISM #formal\_name Mycobacterium leprae  
 DATE 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Sep-1997

ACCESSIONS S72830  
 REFERENCE S72580  
 #authors Smith, D.R.; Robison, K.  
 #submitter submitted to the EMBL Data Library, November 1993  
 #description Mycobacterium leprae cosmid B1620.  
 #accession S72830  
 #status preliminary

```

##molecule-type DNA
##residues 1-87 ##label SMI
##cross-references EMBL:U00015; NID:g466931; PID:g466951
GENETICS
#start-codon GTG
SUMMARY #length 87 #molecular-weight 9272 #checksum 188

Query Match 74.3%; Score 52; DB 2; Length 87;
Best Local Similarity 55.6%; Pred. No. 4.99e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 19 FAMPNFOQL 27
|||:||||
OY 1 FAMPNFOQL 9

RESULT 3
ENTRY C64359 #type complete
TITLE ribosomal protein S5 - Methanococcus jannaschii
ORGANISM #formal name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence-revision 13-Sep-1996 #text-change
10-Oct-1997

ACCESSIONS
REFERENCE C64359
#authors Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kalne, B.P.; Bordovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal Science (1996) 273:1058-1073
#title Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
#cross-references MUID:96337999
#accession C64359
#status preliminary; nucleic acid sequence not shown;
#molecule-type DNA
#residues 1-217 ##label BUL
##cross-references GB:U67497; GB:L77117; NID:g1591160; PID:g1591177;
TIGR:M0475; PID:g1510548

GENETICS
#map-position FOR18436-419089
CLASSIFICATION #superfamily Escherichia coli ribosomal protein S5
SUMMARY #length 217 #molecular-weight 23839 #checksum 5006

Query Match 72.9%; Score 51; DB 2; Length 217;
Best Local Similarity 55.6%; Pred. No. 7.70e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 185 FAMPNFOQL 193
|||:||||
OY 1 FAMPNFOQL 9

RESULT 4
ENTRY J0164 #type fragment
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck hepatitis virus W64 (strain WMS23) (fragment)
ORGANISM #formal name woodchuck hepatitis virus
DATE 31-Mar-1989 #sequence-revision 31-Mar-1989 #text-change
21-Nov-1997

ACCESSIONS
REFERENCE A29498
#authors Ettemble, J.; Moeroy, T.; Trepo, C.; Tiollais, P.; Buendia, M.A.
#journal Gene (1986) 50:207-214
#title Nucleotide sequence of the woodchuck hepatitis virus surface antigen mRNAs and the variability of three overlapping

```

```

viral genes.
#cross-references MUID:87219879
#accession A29498
##molecule-type mRNA
##residues 1-556 ##label ETI
##cross-references GB:M15954; NID:9893289; PID:9336155
GENETICS
#gene P
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase
KEYWORDS DNA biosynthesis; nucleotidyltransferase
SUMMARY #length 556 #checksum 3238

Query Match 72.9%; Score 51; DB 1; Length 556;
Best Local Similarity 77.8%; Pred. No. 7.70e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 120 FAVPNLOQL 128
|||:||||
OY 1 FAMPNFOQL 9

RESULT 5
ENTRY J0164 #type complete
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck hepatitis virus 1
ORGANISM #formal name woodchuck hepatitis virus
DATE 14-Nov-1983 #sequence-revision 14-Nov-1983 #text-change
20-Mar-1998

ACCESSIONS
REFERENCE A00707
#authors Galibert, F.; Chen, T.N.; Mandart, E.
#journal J. Virol. (1984) 41:51-65
#title Nucleotide sequence of a cloned woodchuck hepatitis virus genome: comparison with the hepatitis B virus sequence.
#cross-references MUID:82216969
#accession A00707
##molecule-type DNA
#residues 1-879 ##label GAL
##cross-references GB:J02442; NID:9336126; PID:9336127
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase
KEYWORDS DNA biosynthesis; nucleotidyltransferase
SUMMARY #length 879 #molecular-weight 99185 #checksum 8623

Query Match 72.9%; Score 51; DB 1; Length 879;
Best Local Similarity 77.8%; Pred. No. 7.70e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 443 FAVPNLOQL 451
|||:||||
OY 1 FAMPNFOQL 9

RESULT 6
ENTRY J0164 #type complete
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) - ground squirrel hepatitis virus
ORGANISM #formal name ground squirrel hepatitis virus
DATE 25-Feb-1985 #sequence-revision 25-Feb-1985 #text-change
20-Mar-1998

ACCESSIONS
REFERENCE A00709
#authors Seeger, C.; Ganem, D.; Varmus, H.E.
#journal J. Virol. (1984) 51:367-375
#title Nucleotide sequence of an infectious molecularly cloned genome of ground squirrel hepatitis virus.
#cross-references MUID:84267998
#accession A00709
##molecule-type DNA
#residues 1-881 ##label SEE
##cross-references GB:K02715; NID:9325400; PID:9325402
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase
KEYWORDS DNA biosynthesis; nucleotidyltransferase
SUMMARY #length 881 #molecular-weight 99976 #checksum 6194

```



```

Query Match          72.9%: Score 51; DB 1; Length 881;
Best Local Similarity 77.8%; Pred. No. 7.70e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 445 FAVPNLOTL 453
1 FAMPNFOTL 9

RESULT 7
ENTRY JDLV2 #type complete
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck
        hepatitis virus 2
ORGANISM #formal_name woodchuck hepatitis virus
        #sequence_revision 30-Jun-1987 #text_change
        07-Nov-1997
ACCESSIONS A00708
REFERENCE A03015
#authors Kodama, K.; Ogasawara, N.; Yoshikawa, H.; Murakami, S.
#journal J. Virol. (1985) 56:978-986
#title Nucleotide sequence of a cloned woodchuck hepatitis virus
        genome: evolutionary relationship between hepadnaviruses.
#cross-references MUID:86062931
#accession A00708
##molecule_type DNA
##residues 1-883 #label KOD
#cross-references GB:M1082; NID:9336132; PID:9336134
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase
KEYWORDS DNA biosynthesis; nucleotidyltransferase
SUMMARY #length 883 #molecular_weight 99346 #checksum 593

Query Match          72.9%: Score 51; DB 1; Length 883;
Best Local Similarity 77.8%; Pred. No. 7.70e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 447 FAVPNLOTL 455
1 FAMPNFOTL 9

RESULT 8
ENTRY JDLV59 #type complete
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck
        hepatitis virus 59
ORGANISM #formal_name woodchuck hepatitis virus
        #sequence_revision 30-Jun-1989 #text_change
        14-Nov-1997
ACCESSIONS G29969
REFERENCE A94368
#authors Cohen, J.I.; Miller, R.H.; Rosenblum, B.; Denniston, K.;
        Gerin, J.L.; Purcell, R.H.
#journal Virology (1988) 162:12-20
#title Sequence comparison of woodchuck hepatitis virus replicative
        forms shows conservation of the genome.
#cross-references MUID:88101359
#accession G29969
##molecule_type DNA
##residues 1-884 #label COH
#cross-references GB:M19183; NID:9336141; PID:9336143
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase
KEYWORDS DNA biosynthesis; nucleotidyltransferase
SUMMARY #length 884 #molecular_weight 99399 #checksum 3128

Query Match          72.9%: Score 51; DB 1; Length 884;
Best Local Similarity 77.8%; Pred. No. 7.70e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 448 FAVPNLOTL 456
1 FAMPNFOTL 9

```

```

RESULT 9
ENTRY JDLV8 #type complete
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck
        hepatitis virus 8
ORGANISM #formal_name woodchuck hepatitis virus
        #sequence_revision 31-Mar-1990 #text_change
        25-Oct-1996
ACCESSIONS A32397
REFERENCE A94222
#authors Gitones, R.; Cote, P.J.; Hornbuckle, W.E.; Tennant, B.C.;
        Gerin, J.L.; Purcell, R.H.; Miller, R.H.
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:1846-1849
#title Complete nucleotide sequence of a molecular clone of
        woodchuck hepatitis virus that is infectious in the natural
        host.
#cross-references MUID:89184524
#accession A32397
##molecule_type DNA
##residues 1-884 #label GIR
COMMENT The DNA sequence was obtained from GenBank, release 61.0.
GENETICS
#gene P
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase
KEYWORDS DNA biosynthesis; nucleotidyltransferase
SUMMARY #length 884 #molecular_weight 99708 #checksum 2527

Query Match          72.9%: Score 51; DB 1; Length 884;
Best Local Similarity 77.8%; Pred. No. 7.70e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 448 FAVPNLOTL 456
1 FAMPNFOTL 9

RESULT 10
ENTRY JDLV7 #type complete
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck
        hepatitis virus 7
ORGANISM #formal_name woodchuck hepatitis virus
        #sequence_revision 30-Jun-1989 #text_change
        14-Nov-1997
ACCESSIONS G29969
REFERENCE A94368
#authors Cohen, J.I.; Miller, R.H.; Rosenblum, B.; Denniston, K.;
        Gerin, J.L.; Purcell, R.H.
#journal Virology (1988) 162:12-20
#title Sequence comparison of woodchuck hepatitis virus replicative
        forms shows conservation of the genome.
#cross-references MUID:88101359
#accession G29969
##molecule_type DNA
##residues 1-884 #label COH
#cross-references GB:M18752; NID:9336136; PID:9336138
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase
KEYWORDS DNA biosynthesis; nucleotidyltransferase
SUMMARY #length 884 #molecular_weight 99732 #checksum 4231

Query Match          72.9%: Score 51; DB 1; Length 884;
Best Local Similarity 77.8%; Pred. No. 7.70e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 448 FAVPNLOTL 456
1 FAMPNFOTL 9

RESULT 11
ENTRY B47050 #type complete
TITLE glna 3'-region hypothetical protein - Synechococcus sp.
ALTERNATE_NAMES RPD3/acuc homolog
ORGANISM #formal_name Synechococcus sp.
        #sequence_revision 27-Jan-1995 #text_change
        27-Jan-1995

```

```

REFERENCE
#authors      Mahnke-Zizelman, D.K.; Eddy, R.; Shows, T.B.; Sabina, R.L.
#journal      Blochm. Biophys. Acta (1996) 1306:75-92
#title        Characterization of the human AMPD3 gene reveals that 5' exon
              usage is subject to transcriptional control by three
              tandem promoters and alternative splicing.
#accession    S68147
##status      preliminary
##molecule_type DNA
##residues    1-774 ##label MAH
##cross-references EMBL:U29925
#note         the nucleotide sequence was submitted to the EMBL Data
              library, June 1995
REFERENCE
#note         only a small part of the nucleic acid sequence is shown
#authors      Mahnke-Zizelman, D.K.; Sabina, R.L.
#journal      J. Biol. Chem. (1992) 267:20866-20877
#title        Cloning of human AMP deaminase isoform E cDNAs. Evidence for
              a third AMPD gene exhibiting alternatively spliced
              5'-exons.
#cross-references MVID:93015995
#accession    C45071
##status      preliminary; not compared with conceptual translation
##molecule_type nucleic acid
##residues    1-658 ##label MA2
##cross-references GB:M84722; NID:g178552; PID:g553179
              sequence extracted from NCBI backbone (NCBIPI:116090)
GENERICS
#note         sequence extracted from NCBI backbone (NCBIPI:116090)
#introns
6/1: 81/2; 149/3; 204/1; 277/2; 320/3; 385/3; 429/3; 484/2;
526/3; 581/2; 621/3; 679/3; 716/3
CLASSIFICATION
#superfamily AMP deaminase
#alternative splicing; hydrolase
#length 774 #molecular-weight 89513 #checksum 3787
SUMMARY
Query Match      70.0%; Score 49; DB 2; Length 774;
Best Local Similarity 71.4%; Pred. No. 1.81e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Oy 1 FAMDPQ 7
Db 132 YAMPEQ 138
      |||:|
      1 FAMDPQ 7
RESULT 14
ENTRY      S68146 #type complete
TITLE      AMP deaminase (EC 3.5.4.6) isoform E, splice form Ia - human
ORGANISM   HOMO SAPIENS #common_name man
DATE       06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change
              20-Mar-1998
ACCESSIONS S68146; A45071; B45071; S28149; S27955
REFERENCE
#authors      Mahnke-Zizelman, D.K.; Eddy, R.; Shows, T.B.; Sabina, R.L.
#journal      Biochim. Biophys. Acta (1996) 1306:75-92
#title        Characterization of the human AMPD3 gene reveals that 5' exon
              usage is subject to transcriptional control by three
              tandem promoters and alternative splicing.
#accession    S68146
##status      preliminary
##molecule_type DNA
##residues    1-776 ##label MAH
##cross-references EMBL:U29925
#note         the nucleotide sequence was submitted to the EMBL Data
              library, June 1995
              only a small part of the nucleic acid sequence is shown
              only a small part of the translation is shown
REFERENCE
#note         only a small part of the nucleic acid sequence is shown
#authors      Mahnke-Zizelman, D.K.; Sabina, R.L.
#journal      J. Biol. Chem. (1992) 267:20866-20877
#title        Cloning of human AMP deaminase isoform E cDNAs. Evidence for
              a third AMPD gene exhibiting alternatively spliced
              5'-exons.

```

Search completed: Fri Sep 11 12:54:35 1998  
Job time : 29 secs.

```

#cross-references MUID:93015995
#accession A45071
#molecule-type mRNA
#residues 1-216 #label MA2
#cross-references EMBL:M84720; NID:g178548; PID:g178549
#note splice form 1a (fragment)
#note sequence extracted from NCBI backbone (NCBIP:116076)
#accession B45071
#status preliminary; not compared with conceptual translation
#molecule-type mRNA
#residues 10-776 #label MA3
#cross-references GB:M84721; NID:g178550; PID:g178551
#note sequence extracted from NCBI backbone (NCBIP:116085)
#note splice form 1b; Met-10 is the initiator

REFERENCE
#authors Yamada, Y.; Goto, H.; Ogasawara, N.
#journal Biochim. Biophys. Acta (1992) 1171:125-128
#title Cloning and nucleotide sequence of the cDNA encoding human
        erythrocyte-specific AMP deaminase.
#accession S28149
#molecule-type mRNA
#residues 10-776 #label YAM
#cross-references GB:D12775

GENETICS
#gene GDB:AMPD3
#cross-references GDB:136013; OMIM:102772
#map-position 11p15-11p15
#introns 8/1; 83/2; 151/3; 206/1; 279/2; 322/3; 387/3; 431/3; 486/2;
        528/3; 583/2; 623/3; 681/3; 718/3
CLASSIFICATION #superfamily AMP deaminase
KEYWORDS alternative initiators; alternative splicing; hydrolase
FEATURE
1-776 #product AMP deaminase isoform E, long splice form
        #status predicted #label LSPL\
10-776 #product AMP deaminase isoform E, short splice form
        #status predicted #label SPL
SUMMARY #length 776 #molecular-weight 89727 #checksum 7740

Query Match 70.0%; Score 49; DB 2; Length 776;
Best Local Similarity 71.4%; Pred. No. 1.81e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 134 YAMPEFO 140
OY 1 FAMPNFO 7

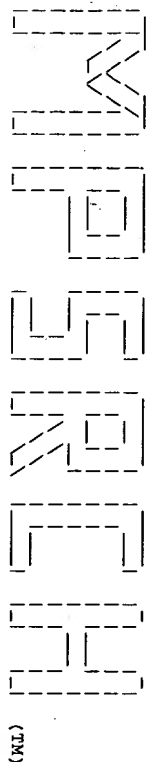
RESULT 15
ENTRY 149364 #type complete
TITLE protein tyrosine phosphatase - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
        02-Jul-1996
ACCESSIONS 149364
REFERENCE 149364
#authors Wisbart, M.J.; Deun, J.M.; Williams, J.A.; Dixon, J.E.
#journal J. Biol. Chem. (1995) 270:26782-26785
#title A single mutation converts a novel-phosphotyrosine binding
        domain into a dual-specificity phosphatase.
#accession 149364
#status preliminary; translated from GB/EMBL/DBD
#molecule-type mRNA
#residues 1-205 #label RBS
#cross-references EMBL:U34973; NID:g1063624; PID:g1063625
SUMMARY #length 205 #molecular-weight 23683 #checksum 2745

Query Match 68.6%; Score 48; DB 2; Length 205;
Best Local Similarity 66.7%; Pred. No. 2.75e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 71 FIKPNFOOL 79
OY 1 FAMPNFO 9

```

**THIS PAGE BLANK (USPTO)**



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Sep 11 12:54:53 1998; MasPar time 2.41 Seconds  
Tabular output not generated. 93.759 Million cell updates/sec

Title: >US-08-452-843-5  
Description: (1-9) from US08452843.pep  
Perfect Score: 70  
Sequence: 1 FAMPNFTL 9

Scoring table:  
PAM 150  
Gap 15

Searched: 69111 seqs, 25083644 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swiss1

Statistics: Mean 24.107; Variance 27.513; scale 0.876

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description                       | Pred. No. |
|------------|-------|-------------|--------|----|-----------------------------------|-----------|
| 1          | 53    | 75.7        | 744    | 1  | BISC_RHOSH BIOTIN SULFOXIDE REDUC | 7.98e-01  |
| 2          | 51    | 72.9        | 217    | 1  | R55_METJA 30S RIBOSOMAL PROTEIN   | 2.18e+00  |
| 3          | 51    | 72.9        | 556    | 1  | DPOL_HPV6 DNA POLYMERASE (EC 2.7  | 2.18e+00  |
| 4          | 51    | 72.9        | 879    | 1  | DPOL_HPV1 DNA POLYMERASE (EC 2.7  | 2.18e+00  |
| 5          | 51    | 72.9        | 881    | 1  | DPOL_HPV8 DNA POLYMERASE (EC 2.7  | 2.18e+00  |
| 6          | 51    | 72.9        | 883    | 1  | DPOL_HPV8 DNA POLYMERASE (EC 2.7  | 2.18e+00  |
| 7          | 51    | 72.9        | 884    | 1  | DPOL_HPV9 DNA POLYMERASE (EC 2.7  | 2.18e+00  |
| 8          | 51    | 72.9        | 884    | 1  | DPOL_HPV9 DNA POLYMERASE (EC 2.7  | 2.18e+00  |
| 9          | 51    | 72.9        | 884    | 1  | DPOL_HPV9 DNA POLYMERASE (EC 2.7  | 2.18e+00  |
| 10         | 51    | 72.9        | 931    | 1  | F480_MOUSE CELL SURFACE GLYCOPRO  | 2.18e+00  |
| 11         | 49    | 70.0        | 310    | 1  | YGLA_SYN2 HYPOTHETICAL 34.1 KD P  | 5.79e+00  |
| 12         | 49    | 70.0        | 456    | 1  | THDF_EUCAP POSSIBLE THIOPEPHE AND | 5.79e+00  |
| 13         | 49    | 70.0        | 767    | 1  | AMP_DEAMINASE 3 (EC 3. 9.34e+00   | 5.79e+00  |
| 14         | 48    | 68.6        | 306    | 1  | Y186_CAEEL DNA POLYMERASE (EC 2.7 | 9.34e+00  |
| 15         | 48    | 68.6        | 481    | 1  | DPOL_HPV2 DNA POLYMERASE (EC 2.7  | 9.34e+00  |
| 16         | 48    | 68.6        | 563    | 1  | HEMA_IAMAA HEMAGGLUTININ PRECURSO | 9.34e+00  |
| 17         | 48    | 68.6        | 564    | 1  | HEMA_IAMAA HEMAGGLUTININ PRECURSO | 9.34e+00  |
| 18         | 48    | 68.6        | 564    | 1  | HEMA_IAMAA HEMAGGLUTININ PRECURSO | 9.34e+00  |
| 19         | 48    | 68.6        | 564    | 1  | HEMA_IAMAA HEMAGGLUTININ PRECURSO | 9.34e+00  |
| 20         | 48    | 68.6        | 564    | 1  | HEMA_IAMAA HEMAGGLUTININ PRECURSO | 9.34e+00  |
| 21         | 48    | 68.6        | 564    | 1  | HEMA_IAMAA HEMAGGLUTININ PRECURSO | 9.34e+00  |
| 22         | 48    | 68.6        | 564    | 1  | HEMA_IAMAA HEMAGGLUTININ PRECURSO | 9.34e+00  |
| 23         | 48    | 68.6        | 568    | 1  | HEMA_IAMAB HEMAGGLUTININ PRECURSO | 9.34e+00  |

|    |    |      |      |   |                                   |          |
|----|----|------|------|---|-----------------------------------|----------|
| 24 | 48 | 68.6 | 657  | 1 | HCVB_PANTIN HEMOCYANIN B CHAIN.   | 9.34e+00 |
| 25 | 48 | 68.6 | 657  | 1 | HCVB_PANTIN HEMOCYANIN A CHAIN.   | 9.34e+00 |
| 26 | 48 | 68.6 | 730  | 1 | DPOL_HPV4 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 27 | 48 | 68.6 | 750  | 1 | DPOL_HPV2 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 28 | 48 | 68.6 | 763  | 1 | DPOL_HPV2 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 29 | 48 | 68.6 | 832  | 1 | DPOL_HPV4 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 30 | 48 | 68.6 | 832  | 1 | DPOL_HPV4 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 31 | 48 | 68.6 | 832  | 1 | DPOL_HPV4 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 32 | 48 | 68.6 | 832  | 1 | DPOL_HPV4 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 33 | 48 | 68.6 | 843  | 1 | DPOL_HPV4 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 34 | 48 | 68.6 | 843  | 1 | DPOL_HPV4 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 35 | 48 | 68.6 | 843  | 1 | DPOL_HPV4 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 36 | 48 | 68.6 | 843  | 1 | DPOL_HPV4 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 37 | 48 | 68.6 | 843  | 1 | DPOL_HPV4 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 38 | 48 | 68.6 | 843  | 1 | DPOL_HPV4 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 39 | 48 | 68.6 | 845  | 1 | DPOL_HPV2 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 40 | 48 | 68.6 | 845  | 1 | DPOL_HPV2 DNA POLYMERASE (EC 2.7  | 9.34e+00 |
| 41 | 48 | 68.6 | 959  | 1 | YV34_MYCLE HYPOTHETICAL 105.2 KD  | 9.34e+00 |
| 42 | 48 | 68.6 | 2476 | 1 | ZAN_PIG ZONADHESIN PRECURSOR.     | 9.34e+00 |
| 43 | 47 | 67.1 | 578  | 1 | HEMA_INBME HEMAGGLUTININ PRECURSO | 1.49e+01 |
| 44 | 47 | 67.1 | 583  | 1 | HEMA_INBSE HEMAGGLUTININ PRECURSO | 1.49e+01 |
| 45 | 47 | 67.1 | 994  | 1 | PPOL_DROME POLY (ADP-RIBOSE) POLY | 1.49e+01 |

# ALIGNMENTS

| RESULT ID  | 1   | BISC_RHOSH | STANDARD; | PRT; | 744 AA. |
|--|---|------------|-----------|------|---------|
| AC   | P54934;   |            |           |      |         |
| DT   | 01-OCT-1996 (REL. 34, CREATED)  |            |           |      |         |
| DT   | 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)   |            |           |      |         |
| DT   | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)   |            |           |      |         |
| DE   | BIOTIN SULFOXIDE REDUCTASE (EC 1.-.-.-) (BDS REDUCTASE) (BSO REDUCTASE).  |            |           |      |         |
| OS   | RHODOACTER SPHEROIDES (RHODOSEUDOMONAS SPHEROIDES).   |            |           |      |         |
| OC   | PROKARYOTA: GRACILICUTES; ANOXIPHOTOBACTERIA; PURPLE BACTERIA; RHODOSPIRILLACEAE.   |            |           |      |         |
| CC   | (1)   |            |           |      |         |
| CC   | SEQUENCE FROM N.A.  |            |           |      |         |
| CC   | STRAIN-SP. DENTRIFICANS IL106;  |            |           |      |         |
| CC   | MEDLINE: 95251380.  |            |           |      |         |
| CC   | POLLOCK V.V., BARBER M.J.:  |            |           |      |         |
| CC   | ARCH. BIOCHEM. BIOPHYS. 318:322-332(1995).  |            |           |      |         |
| CC   | -1- FUNCTION: THIS ENZYME MAY SERVE AS A SCAVENGER, ALLOWING THE CELL TO UTILIZE BIOTIN SULFOXIDE AS A BIOTIN SOURCE (BY SIMILARITY). |            |           |      |         |
| CC   | -1- CATALYTIC ACTIVITY: REDUCES A SPONTANEOUS OXIDATION PRODUCT OF BIOTIN, D-BIOTIN D-SULFOXIDE (BSO OR BDS), BACK TO BIOTIN.         |            |           |      |         |
| CC   | -1- COFACTOR: MOLYBDENUM (MOLYBDOPTERIN).   |            |           |      |         |
| CC   | -1- SIMILARITY: TO OTHER PROKARYOTIC MOLYBDOPTERIN-CONTAINING OXIDOREDUCTASES.  |            |           |      |         |
| CC   | EMBL: U08189; G953224; -.   |            |           |      |         |
| DR   | PROSITE: PS00551; MOLYBDOPTERIN_PROK_1; FALSE_NEG.  |            |           |      |         |
| DR   | PROSITE: PS00490; MOLYBDOPTERIN_PROK_2; 1.  |            |           |      |         |
| DR   | PROSITE: PS00932; MOLYBDOPTERIN_PROK_3; FALSE_NEG.  |            |           |      |         |
| KW   | OXIDOREDUCTASE; MOLYBDENUM.   |            |           |      |         |
| SC   | SEQUENCE 744 AA; 80266 MW; 6B6E3E56 CRC32;  |            |           |      |         |
| Query Match  |   |            |           |      |         |
| Best Local Similarity 75.7%; Score 53; DB 1; Length 744;   |   |            |           |      |         |
| Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0; |   |            |           |      |         |
| DB   | 519 FEMPEPTEF 527   |            |           |      |         |
| Oy   | 1 FAMPNFTL 9  |            |           |      |         |
| RESULT ID  | 2   | R55_METJA  | STANDARD; | PRT; | 217 AA. |
| AC   | P54045;   |            |           |      |         |
| DT   | 01-OCT-1996 (REL. 34, CREATED)  |            |           |      |         |
| DT   | 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)   |            |           |      |         |
| DT   | 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)   |            |           |      |         |
| DE   | 30S RIBOSOMAL PROTEIN S5P.  |            |           |      |         |

GN M00475.  
OS METHANOCOCCUS JANNASCHII.  
OC ARCHAEABACTERIA: EURYARCHAEOTA: METHANOCOCCALES; METHANOCOCCACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96337999.  
RA BOLT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMH J.-F., ADAMS M.D., REICH C.I.,  
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,  
RA SCOTT J.L., GEORGHEN N.S.M., WEIDMAN J.F., FUHRMAN J.L., NGUYEN D.,  
RA COTTON M.D., ROBERTS K.M., HURST M.A., KATNE B.P., BORDOVSKY M.,  
RA KLEIN H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.,  
RL SCIENCE 273:1058-1073(1996).  
CC -1- SIMILARITY: BELONGS TO THE SSP FAMILY OF RIBOSOMAL PROTEINS.  
DR EMBL: U67497, G1591177.  
DR PROSITE: P500585; RIBOSOMAL\_S5; 1.  
DR TIGR: M00475.  
KW RIBOSOMAL PROTEIN.  
SQ SEQUENCE 217 AA: 23839 MW; 4D36A4B3 CRC32;  
  
Query Match 72.9%; Score 51; DB 1; Length 217;  
Best Local Similarity 55.6%; Pred. No. 2.18e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
Db 185 FAMAFPEAL 193  
|||:|:|  
OY 1 FAMAFPEAL 9  
  
RESULT 3  
ID DPOL\_MHY6 STANDARD; PRT; 556 AA.  
AC P11292;  
DT 01-JUL-1989 (REL. 11, CREATED)  
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)  
DT 01-JUL-1989 (REL. 11, LAST ANNOTATION UPDATE)  
DE DNA POLYMERASE (EC 2.7.7.7) (FRAGMENT).  
GN P.  
OS WOODCHUCK HEPATITIS VIRUS W64 (ISOLATE PMS23).  
OC VIRIDAE: DS-DNA ENVELOPED VIRUSES; HEPADNAVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87219879.  
RA ETIEMBLE J., MOERLEY T., TREPO C., TIOLLAIS P., BUENDIA M.-A.;  
RL GENE 50:207-214(1986).  
DR EMBL: M15954; G336135; -.  
DR PIR: A29498; J0VL64.  
KW DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION.  
FT NON\_TER 1  
SQ SEQUENCE 556 AA: 61871 MW; D64F0695 CRC32;  
  
Query Match 72.9%; Score 51; DB 1; Length 556;  
Best Local Similarity 77.8%; Pred. No. 2.18e+00;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 120 FAVPNLOTL 128  
|||:|:|  
OY 1 FAVPNLOTL 9  
  
RESULT 4  
ID DPOL\_MHY1 STANDARD; PRT; 879 AA.  
AC P03160;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
DT 01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)  
DE DNA POLYMERASE (EC 2.7.7.7).  
GN P.  
OS WOODCHUCK HEPATITIS VIRUS 1.  
OC VIRIDAE: DS-DNA ENVELOPED VIRUSES; HEPADNAVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE: 82216969.  
RA GALIBERT F., CHEN T.N., MANDART E.;  
RL J. VIROL. 41:51-65(1982).  
DR EMBL: J02442; E29063; ALT\_SEQ.  
DR PIR: A00707; J0VL6.  
KW DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION.  
SQ SEQUENCE 879 AA: 99185 MW; 3BD450AF CRC32;  
  
Query Match 72.9%; Score 51; DB 1; Length 879;  
Best Local Similarity 77.8%; Pred. No. 2.18e+00;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 443 FAVPNLOTL 451  
|||:|:|  
OY 1 FAVPNLOTL 9  
  
RESULT 5  
ID DPOL\_HPBGS STANDARD; PRT; 881 AA.  
AC P03161;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)  
DE DNA POLYMERASE (EC 2.7.7.7) (A PROTEIN).  
GN P.  
OS GROUND SQUIRREL HEPATITIS VIRUS (GSHV).  
OC VIRIDAE: DS-DNA ENVELOPED VIRUSES; HEPADNAVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 84267998.  
RA SEEGER C., GANEM D., VARBUS H.E.;  
RL J. VIROL. 51:367-375(1984).  
DR EMBL: K02715; G325402; -.  
DR PIR: A00709; J0VL6.  
KW DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION.  
SQ SEQUENCE 881 AA: 99976 MW; 2295D041 CRC32;  
  
Query Match 72.9%; Score 51; DB 1; Length 881;  
Best Local Similarity 77.8%; Pred. No. 2.18e+00;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 445 FAVPNLOTL 453  
|||:|:|  
OY 1 FAVPNLOTL 9  
  
RESULT 6  
ID DPOL\_MHY8 STANDARD; PRT; 883 AA.  
AC P06275;  
DT 01-JAN-1988 (REL. 06, CREATED)  
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)  
DT 01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)  
DE DNA POLYMERASE (EC 2.7.7.7).  
GN P.  
OS WOODCHUCK HEPATITIS VIRUS 8.  
OC VIRIDAE: DS-DNA ENVELOPED VIRUSES; HEPADNAVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 86062931.  
RA KODAMA K., OGASAWARA N., YOSHIKAWA H., MURAKAMI S.;  
RL J. VIROL. 56:978-986(1986).  
DR EMBL: M11082; G336134; -.  
DR PIR: A00708; J0VL62.  
KW DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION.  
SQ SEQUENCE 883 AA: 99346 MW; 843B7F01 CRC32;  
  
Query Match 72.9%; Score 51; DB 1; Length 883;  
Best Local Similarity 77.8%; Pred. No. 2.18e+00;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 447 FAVPNLOTL 455  
|||:|:|  
OY 1 FAVPNLOTL 9

| Query                 | Match   | Score                             | DB 1  | Length                           | 884   |
|-----------------------|---|-----------------------------------|-------|----------------------------------|-------|
| Best Local Similarity | 77.8%   | 72.9%                             | 77.8% | 72.9%                            | 72.9% |
| Matches               | 7   | Conservative                      | 2     | Mismatches                       | 0     |
|                       |   |                                   |       | Indels                           | 0     |
|                       |   |                                   |       | Gaps                             | 0     |
| DB                    | 448   | FAYNDLOTL                         | 456   |                                  |       |
| QY                    | 1   | FAMNPFQTL                         | 9     |                                  |       |
|                       |   |                                   |       |                                  |       |
| RESULT                | 10  |                                   |       |                                  |       |
| ID                    | F480_MOUSE  | STANDARD                          | PRT   | 931                              | AA    |
| AC                    | 061549  |                                   |       |                                  |       |
| DT                    | 01-NOV-1997   | (REL. 35, CREATED)                |       |                                  |       |
| DT                    | 01-NOV-1997   | (REL. 35, LAST SEQUENCE UPDATE)   |       |                                  |       |
| DT                    | 01-NOV-1997   | (REL. 35, LAST ANNOTATION UPDATE) |       |                                  |       |
| DE                    | CELL SURFACE GLYCOPROTEIN F4/80 PRECURSOR.                          |                                   |       |                                  |       |
| GN                    | GPF480.   |                                   |       |                                  |       |
| OS                    | MUS MUSCULUS (MOUSE).   |                                   |       |                                  |       |
| OC                    | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;      |                                   |       |                                  |       |
| OC                    | EUTHERIA; RODENTIA.   |                                   |       |                                  |       |
| RN                    | [1]   |                                   |       |                                  |       |
| RP                    | SEQUENCE FROM N.A.  |                                   |       |                                  |       |
| RC                    | STRAIN-BALB/C; TISSUE-PERITONEAL CAVITY;                            |                                   |       |                                  |       |
| RX                    | MEDLINE; 96132946.  |                                   |       |                                  |       |
| RA                    | MCKNIGHT A.J., MACFARLANE A.J., DRI P., TURLEY L., WILLIS A.C.,     |                                   |       |                                  |       |
| RA                    | GORDON S.;  |                                   |       |                                  |       |
| RL                    | J. BIOL. CHEM. 271:486-489(1996).                                   |                                   |       |                                  |       |
| CC                    | -1- FUNCTION: PROBABLY INVOLVED IN CELL ADHESION WITHIN TISSUES     |                                   |       |                                  |       |
| CC                    | AND RECEPTOR SIGNALING.   |                                   |       |                                  |       |
| CC                    | -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.                |                                   |       |                                  |       |
| CC                    | -1- TISSUE SPECIFICITY: IN MACROPHAGES; BUT ABSENT FROM THOSE WHICH |                                   |       |                                  |       |
| CC                    | ARE LOCALIZED WITHIN T-CELL AREAS OF LYMPH NODES AND SPLEEN.        |                                   |       |                                  |       |
| CC                    | LOW LEVEL OF EXPRESSION ON BLOOD MONOCYTES.                         |                                   |       |                                  |       |
| CC                    | -1- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.                        |                                   |       |                                  |       |
| CC                    | -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS. |                                   |       |                                  |       |
| DR                    | EMBL; X93328; E214264; -.   |                                   |       |                                  |       |
| DR                    | MGI; MGI:105054; GPF480.  |                                   |       |                                  |       |
| DR                    | PROSITE; PS00010; ASX_HYDROXYL_6                                    |                                   |       |                                  |       |
| DR                    | PROSITE; PS00649; G-PROTEIN_RECP_F2_1; FALSE_NEG.                   |                                   |       |                                  |       |
| DR                    | PROSITE; PS00650; G-PROTEIN_RECP_F2_2; 1.                           |                                   |       |                                  |       |
| DR                    | PROSITE; PS01186; EGF_2; 1.   |                                   |       |                                  |       |
| DR                    | PROSITE; PS01187; EGF_CA_5.   |                                   |       |                                  |       |
| KW                    | G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; RECEPTOR; GLYCOPROTEIN;  |                                   |       |                                  |       |
| KW                    | EGF-LIKE DOMAIN; REPEAT; SIGNAL.                                    |                                   |       |                                  |       |
| FT                    | CHAIN   | 1                                 | 27    | POTENTIAL.                       |       |
| FT                    | CHAIN   | 28                                | 931   | CELL SURFACE GLYCOPROTEIN F4/80. |       |
| FT                    | DOMAIN  | 28                                | 644   | EXTRACELLULAR (POTENTIAL).       |       |
| FT                    | TRANSMEM  | 645                               | 672   | POTENTIAL.                       |       |
| FT                    | DOMAIN  | 673                               | 679   | CYTOPLASMIC (POTENTIAL).         |       |
| FT                    | TRANSMEM  | 680                               | 701   | POTENTIAL.                       |       |
| FT                    | DOMAIN  | 702                               | 711   | EXTRACELLULAR (POTENTIAL).       |       |
| FT                    | TRANSMEM  | 712                               | 735   | POTENTIAL.                       |       |
| FT                    | DOMAIN  | 736                               | 754   | CYTOPLASMIC (POTENTIAL).         |       |
| FT                    | TRANSMEM  | 755                               | 776   | POTENTIAL.                       |       |
| FT                    | DOMAIN  | 777                               | 792   | EXTRACELLULAR (POTENTIAL).       |       |
| FT                    | TRANSMEM  | 793                               | 821   | POTENTIAL.                       |       |
| FT                    | DOMAIN  | 822                               | 839   | CYTOPLASMIC (POTENTIAL).         |       |
| FT                    | TRANSMEM  | 840                               | 859   | POTENTIAL.                       |       |
| FT                    | DOMAIN  | 860                               | 874   | EXTRACELLULAR (POTENTIAL).       |       |
| FT                    | TRANSMEM  | 875                               | 897   | POTENTIAL.                       |       |
| FT                    | DOMAIN  | 898                               | 931   | CYTOPLASMIC (POTENTIAL).         |       |
| FT                    | DOMAIN  | 32                                | 80    | EGF-LIKE 1.                      |       |
| FT                    | DOMAIN  | 81                                | 132   | EGF-LIKE 2.                      |       |

Query Match. 72.9% Score 51; DB 1; Length 931;  
Best Local Similarity 55.6%; Pred. No. 2.18e+00;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

| ID | YGLA_SYN2   | STANDARD | PRT | 310 AA |
|----|---|----------|-----|--------|
| AC | P28606  |          |     |        |
| DT | 01-DEC-1992 (REL. 24, CREATED)  |          |     |        |
| DT | 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)                           |          |     |        |
| DT | 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)                         |          |     |        |
| DE | HYPOTHEICAL 34.1 KD PROTEIN IN GLNA 3 REGION.                         |          |     |        |
| OS | SYNECHOCOCUS SP. (STRAIN PCC 7002) (AGMENELLUM QUADRUPPLICATUM).      |          |     |        |
| OC | PROKARYOTA: GRACILICUTES: OXYPHOTOBACTERIA:                           |          |     |        |
| CC | CYANOBACTERIA (BLUE-GREEN ALGAE); CHROCOCCALES.                       |          |     |        |
| CC | (1)   |          |     |        |
| RP | SEQUENCE FROM N.A.  |          |     |        |
| RC | STRAIN-PR-6;  |          |     |        |
| RC | MEDLINE: 93139025.  |          |     |        |
| RA | WAGNER S.J., THOMAS S.P., KAUFMAN R.I., NIXON B.T., STEVENS S.E. JR.; |          |     |        |
| RL | J. BACTERIOL. 175:604-612(1993).                                      |          |     |        |
| DR | EMBL: Z13965; G380726; -  |          |     |        |
| KW | PIR: S23853; S23853.  |          |     |        |
| SO | SEQUENCE 310 AA; 34145 MW; 808212D2 CRC32;                            |          |     |        |

Query Match. 70.0% Score 49; DB 1; Length 310;  
Best Local Similarity 55.6%; Pred. No. 5.79e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

| ID | YGLA_SYN2   | STANDARD | PRT | 310 AA |
|----|---|----------|-----|--------|
| AC | P28606  |          |     |        |
| DT | 01-DEC-1992 (REL. 24, CREATED)  |          |     |        |
| DT | 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)                           |          |     |        |
| DT | 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)                         |          |     |        |
| DE | HYPOTHEICAL 34.1 KD PROTEIN IN GLNA 3 REGION.                         |          |     |        |
| OS | SYNECHOCOCUS SP. (STRAIN PCC 7002) (AGMENELLUM QUADRUPPLICATUM).      |          |     |        |
| OC | PROKARYOTA: GRACILICUTES: OXYPHOTOBACTERIA:                           |          |     |        |
| CC | CYANOBACTERIA (BLUE-GREEN ALGAE); CHROCOCCALES.                       |          |     |        |
| CC | (1)   |          |     |        |
| RP | SEQUENCE FROM N.A.  |          |     |        |
| RC | STRAIN-PR-6;  |          |     |        |
| RC | MEDLINE: 93139025.  |          |     |        |
| RA | WAGNER S.J., THOMAS S.P., KAUFMAN R.I., NIXON B.T., STEVENS S.E. JR.; |          |     |        |
| RL | J. BACTERIOL. 175:604-612(1993).                                      |          |     |        |
| DR | EMBL: Z13965; G380726; -  |          |     |        |
| KW | PIR: S23853; S23853.  |          |     |        |
| SO | SEQUENCE 310 AA; 34145 MW; 808212D2 CRC32;                            |          |     |        |

Query Match. 70.0% Score 49; DB 1; Length 456;  
Best Local Similarity 55.6%; Pred. No. 5.79e+00;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

| ID | YGLA_SYN2   | STANDARD | PRT | 456 AA |
|----|---|----------|-----|--------|
| AC | P28606  |          |     |        |
| DT | 01-DEC-1992 (REL. 24, CREATED)  |          |     |        |
| DT | 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)                           |          |     |        |
| DT | 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)                         |          |     |        |
| DE | HYPOTHEICAL 34.1 KD PROTEIN IN GLNA 3 REGION.                         |          |     |        |
| OS | SYNECHOCOCUS SP. (STRAIN PCC 7002) (AGMENELLUM QUADRUPPLICATUM).      |          |     |        |
| OC | PROKARYOTA: GRACILICUTES: OXYPHOTOBACTERIA:                           |          |     |        |
| CC | CYANOBACTERIA (BLUE-GREEN ALGAE); CHROCOCCALES.                       |          |     |        |
| CC | (1)   |          |     |        |
| RP | SEQUENCE FROM N.A.  |          |     |        |
| RC | STRAIN-PR-6;  |          |     |        |
| RC | MEDLINE: 93139025.  |          |     |        |
| RA | WAGNER S.J., THOMAS S.P., KAUFMAN R.I., NIXON B.T., STEVENS S.E. JR.; |          |     |        |
| RL | J. BACTERIOL. 175:604-612(1993).                                      |          |     |        |
| DR | EMBL: Z13965; G380726; -  |          |     |        |
| KW | PIR: S23853; S23853.  |          |     |        |
| SO | SEQUENCE 310 AA; 34145 MW; 808212D2 CRC32;                            |          |     |        |

Query Match. 70.0% Score 49; DB 1; Length 456;  
Best Local Similarity 55.6%; Pred. No. 5.79e+00;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

| ID | YGLA_SYN2   | STANDARD | PRT | 456 AA |
|----|---|----------|-----|--------|
| AC | P28606  |          |     |        |
| DT | 01-DEC-1992 (REL. 24, CREATED)  |          |     |        |
| DT | 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)                           |          |     |        |
| DT | 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)                         |          |     |        |
| DE | HYPOTHEICAL 34.1 KD PROTEIN IN GLNA 3 REGION.                         |          |     |        |
| OS | SYNECHOCOCUS SP. (STRAIN PCC 7002) (AGMENELLUM QUADRUPPLICATUM).      |          |     |        |
| OC | PROKARYOTA: GRACILICUTES: OXYPHOTOBACTERIA:                           |          |     |        |
| CC | CYANOBACTERIA (BLUE-GREEN ALGAE); CHROCOCCALES.                       |          |     |        |
| CC | (1)   |          |     |        |
| RP | SEQUENCE FROM N.A.  |          |     |        |
| RC | STRAIN-PR-6;  |          |     |        |
| RC | MEDLINE: 93139025.  |          |     |        |
| RA | WAGNER S.J., THOMAS S.P., KAUFMAN R.I., NIXON B.T., STEVENS S.E. JR.; |          |     |        |
| RL | J. BACTERIOL. 175:604-612(1993).                                      |          |     |        |
| DR | EMBL: Z13965; G380726; -  |          |     |        |
| KW | PIR: S23853; S23853.  |          |     |        |
| SO | SEQUENCE 310 AA; 34145 MW; 808212D2 CRC32;                            |          |     |        |

Query Match. 70.0% Score 49; DB 1; Length 456;  
Best Local Similarity 55.6%; Pred. No. 5.79e+00;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

| ID | YGLA_SYN2   | STANDARD | PRT | 456 AA |
|----|---|----------|-----|--------|
| AC | P28606  |          |     |        |
| DT | 01-DEC-1992 (REL. 24, CREATED)  |          |     |        |
| DT | 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)                           |          |     |        |
| DT | 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)                         |          |     |        |
| DE | HYPOTHEICAL 34.1 KD PROTEIN IN GLNA 3 REGION.                         |          |     |        |
| OS | SYNECHOCOCUS SP. (STRAIN PCC 7002) (AGMENELLUM QUADRUPPLICATUM).      |          |     |        |
| OC | PROKARYOTA: GRACILICUTES: OXYPHOTOBACTERIA:                           |          |     |        |
| CC | CYANOBACTERIA (BLUE-GREEN ALGAE); CHROCOCCALES.                       |          |     |        |
| CC | (1)   |          |     |        |
| RP | SEQUENCE FROM N.A.  |          |     |        |
| RC | STRAIN-PR-6;  |          |     |        |
| RC | MEDLINE: 93139025.  |          |     |        |
| RA | WAGNER S.J., THOMAS S.P., KAUFMAN R.I., NIXON B.T., STEVENS S.E. JR.; |          |     |        |
| RL | J. BACTERIOL. 175:604-612(1993).                                      |          |     |        |
| DR | EMBL: Z13965; G380726; -  |          |     |        |
| KW | PIR: S23853; S23853.  |          |     |        |
| SO | SEQUENCE 310 AA; 34145 MW; 808212D2 CRC32;                            |          |     |        |





**THIS PAGE BLANK (USPTO)**



RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DUREIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON I.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALLON N., SMITH A., SONNHAMER E., STADEN R., SULSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
 RA NATURE 368:32-38(1994).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA DANTE M., KRAMER J., TWYMAN B.,  
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WATERSTON R.,  
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AF016668; G2315648; -  
 SQ SEQUENCE 490 AA; 56299 MW; 823A74E0 CRC32;

Query Match 77.1%; Score 54; DB 3; Length 490;  
 Best Local Similarity 66.7%; Pred. No. 1.35e+00;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 359 FLMKFOVL 367  
 1 1 1 1 1 1  
 QY 1 FAMNFOVL 9

RESULT 3  
 ID 049735; PRELIMINARY; PRT; 87 AA.  
 AC 049735;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 9.3 KD PROTEIN B1620\_F1.14.  
 GN B1620\_F1.14.  
 OS MYCOBACTERIUM LEPRAE.  
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SMITH D.R., ROBISON K.,  
 RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U00015; G466951; -  
 KW HYPOTHETICAL PROTEIN.  
 FT NON\_TER  
 SQ SEQUENCE 87 AA; 9272 MW; 6AF56082 CRC32;

Query Match 74.3%; Score 52; DB 9; Length 87;  
 Best Local Similarity 55.6%; Pred. No. 3.43e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 19 FGMTNFOAM 27  
 1 1 1 1 1 1  
 QY 1 FAMNFOVL 9

RESULT 4  
 ID 022215; PRELIMINARY; PRT; 262 AA.  
 AC 022215;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE T05B9.1  
 OS CAENORHABDITIS ELIGANS.  
 OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SWINBURNE J.,

RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J.,  
 RA COULSON A., CRAXTON M., DEAR S., DU Z., DUREIN R., FAVELLO A.,  
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
 RA JOHNSTON I., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
 RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,  
 RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFKEN L., ROOPRA A.,  
 RA SAUNDERS D., SHOWNKEEN R., SMALLON N., SMITH A., SONNHAMER E.,  
 RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
 RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
 RA WILKINSON-SPROAT J., WOHLDMAN P.,  
 RL NATURE 368:32-38(1994).  
 DR EMBL: 249129; G790395; -  
 SQ SEQUENCE 262 AA; 30162 MW; D3F796CF CRC32;

Query Match 74.3%; Score 52; DB 3; Length 262;  
 Best Local Similarity 55.6%; Pred. No. 3.43e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 79 FOMAFOSL 87  
 1 1 1 1 1 1  
 QY 1 FAMNFOVL 9

RESULT 5  
 ID 008745; PRELIMINARY; PRT; 255 AA.  
 AC 008745;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE EGF-LIKE MODULE CONTAINING, MUCIN-LIKE, HORMONE RECEPTOR-LIKE SEQUENCE  
 DE 1 (EMRL) (FRAGMENT).  
 GN EMRL  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA LIN H.H., STUBBS L.J., MUCENSKI M.L.,  
 RL GENOMICS 41:301-308(1997).  
 DR EMBL: U66892; G2078512; -  
 DR MGD; MGI:106912; EMRL  
 DR PROSITE; PS00010; ASX-HYDROXYL; 4.  
 DR PROSITE; PS01187; EGF-CA; 4.  
 KW GLYCOPROTEIN; EGF-LIKE DOMAIN.  
 FT NON\_TER  
 SQ SEQUENCE 255 AA; 27764 MW; 0FA23F4D CRC32;

Query Match 72.9%; Score 51; DB 10; Length 255;  
 Best Local Similarity 55.6%; Pred. No. 5.42e+00;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 113 FSLPTFOIL 121  
 1 1 1 1 1 1  
 QY 1 FAMNFOVL 9

RESULT 6  
 ID 008744; PRELIMINARY; PRT; 304 AA.  
 AC 008744;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE EGF-LIKE MODULE CONTAINING, MUCIN-LIKE, HORMONE RECEPTOR-LIKE SEQUENCE  
 DE 1 (EMRL) (FRAGMENT).  
 GN EMRL  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.

```

RN [1]
RP SEQUENCE FROM N.A.
RA LIN H.H., STUBBS L.J., MCCENSKI M.L.;
RL GENOMICS 41:301-308(1997);
DR EMBL: U66891; G2078510;
DR MGD: MGI:106912; EMRL.
DR PROSITE: PS00010; ASX_HYDROXYL; 5.
DR PROSITE: PS01187; EGF_CA; 4.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT NON_TER 304
SQ SEQUENCE 304 AA; 33013 MW; F967B4BF CRC32;

Query Match
Best Local Similarity 72.9%; Score 51; DB 10; Length 304;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 162 FSLPTFOIL 170
QY 1 FAMPNFOTL 9

RESULT 7
ID 083759 PRELIMINARY; PRT; 427 AA.
AC 083759;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE ENVELOPE PROTEIN PRECURSOR.
OS WOODCHUCK HEPATITIS VIRUS.
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HEPADNAVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RA GALBERT F., CHEN T.N., MANDART E.;
RL J. VIROL. 41:51-65(1982).
DR EMBL: J02442; E29064;
KW ENVELOPE PROTEIN.
FT CHAIN 99
SQ SEQUENCE 427 AA; 47787 MW; CDC9EF70 CRC32;

Query Match
Best Local Similarity 72.9%; Score 51; DB 11; Length 427;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 149 FAVPNLOTL 157
QY 1 FAMPNFOTL 9

RESULT 8
ID 089244 PRELIMINARY; PRT; 585 AA.
AC 089244;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE POLYMERASE PROTEIN (FRAGMENT).
OS WOODCHUCK HEPATITIS VIRUS.
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HEPADNAVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-TOMPKINS COUNTY, N.Y.;
RX MEDLINE: 93255897
RA KEW M.C., TENNANT B.C., PORCELL R.H., MILLER R.H.;
RL VIRUS RES. 27:229-237(1993).
DR EMBL: M90520; G336159;
FT NON_TER 1
SQ SEQUENCE 585 AA; 65216 MW; FFE6E996 CRC32;

Query Match
Best Local Similarity 72.9%; Score 51; DB 11; Length 585;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 149 FAVPNLOTL 157

```

```

QY 1 FAMPNFOTL 9

RESULT 9
ID P87544 PRELIMINARY; PRT; 931 AA.
AC P87544;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE 104X PROTEIN.
OS BEET SOIL-BORNE VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FURUVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-AHLON;
RA KOENIG R., COMMANDEUR U., LOSS S., BEIER C., KAUFMANN A.,
RA LESEMAN D.E.;
RL J. GEN. VIROL. 78:469-477(1997).
DR EMBL: U64512; G1841519;
SQ SEQUENCE 931 AA; 103802 MW; 92A51B31 CRC32;

Query Match
Best Local Similarity 72.9%; Score 51; DB 11; Length 931;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 586 FAMPNFOTL 594
QY 1 FAMPNFOTL 9

RESULT 10
ID 004459 PRELIMINARY; PRT; 449 AA.
AC 004459;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)
DE F2159.21.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
RN [1]
RP SEQUENCE FROM N.A.
RA DEMAR K., BOEHLER E., FENG J., KIM C., LI Y., SHINN P., SUN H.,
RA CONWAY A., CONWAY A., KURTZ D., OJI O., OSBORNE B., SHEN Y.K.,
RA TORIUMI M., VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A.,
RA THEOLOGIS A., ECKER J.R.;
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AC000103; G2213629;
SQ SEQUENCE 449 AA; 50062 MW; E4E93445 CRC32;

Query Match
Best Local Similarity 71.4%; Score 50; DB 8; Length 449;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 328 FAMPNFOTL 336
QY 1 FAMPNFOTL 9

RESULT 11
ID 024135 PRELIMINARY; PRT; 2354 AA.
AC 024135;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE MET-41.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-1-54.2, 14C4-6;

```

RA MEDLINE: 95401271.  
 RA HART K.L., SANTERRE A., SEKELSKY J.J., MCKIM K.S., BOYD J.B.,  
 RA HANLEY R.S.,  
 RL CELL 82:815-821(1995).  
 DR EMBL: U34925; G998353;  
 DR FLYBASE: FBgn0004367; mei-41.  
 SQ SEQUENCE 2354 AA; 270445 MW; 1E0FB72 CRC32;

Query Match 71.4%; Score 50; DB 3; Length 2354;  
 Best Local Similarity 55.6%; Pred. No. 8.51e+00;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 764 FVMSFOSL 772  
 QY 1 FAMPNFOVL 9

RESULT 12 PRELIMINARY; PRT; 180 AA.

ID 022587  
 AC 022587;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE COSMID T19D7.  
 GN T19D7.6.  
 OS CAENORHABDITIS ELEGANS.  
 CC EURAROTIA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
 RN [1]

SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL NZ;  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONEFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J.,  
 RA COULSON A., CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,  
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLER L., JIER M.,  
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
 RA LARREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,  
 RA O'CALLAGHAN M., PARSONS J., PERCY C., RIEKEN L., ROOPRA A.,  
 RA SANDERS D., SHOWNKEN R., SMALDON N., SMITH A., SONNHAMER E.,  
 RA STADEN R., STOLSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,  
 RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
 RA WILKINSON-SPROAT J., WOHLIDMAN P.;  
 RL NATURE 368:32-38(1994).  
 RN [2]

SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL NZ;  
 RA MIX P.;  
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]

SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL NZ;  
 RA WATERSTON R.;  
 RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U56961; G1293810;  
 SQ SEQUENCE 180 AA; 20433 MW; C1203EB9 CRC32;

Query Match 70.0%; Score 49; DB 3; Length 180;  
 Best Local Similarity 55.6%; Pred. No. 1.33e+01;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 87 FMPKFRNL 95  
 QY 1 FAMPNFOVL 9

RESULT 13 PRELIMINARY; PRT; 186 AA.

ID 006450  
 AC 006450;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)  
 DE DNAG, RPOD, CPOA GENES AND ORF3 AND ORF5  
 DE (FRAGMENT).

GN DNAG.  
 OS STREPTOCOCCUS PNEUMONIAE.  
 CC PROKARYOTA; FIRMICUTES; COCCI; STREPTOCOCCAEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-R6;  
 RA GREBE T., PAIK J., HAKENBECK R.;  
 RL J. BACTERIOL. 179:3342-3349(1997).  
 DR EMBL: Y11463; E316588;  
 FT NON\_TER  
 SQ SEQUENCE 186 AA; 21698 MW; 378F26AC CRC32;

Query Match 70.0%; Score 49; DB 9; Length 186;  
 Best Local Similarity 66.7%; Pred. No. 1.33e+01;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 85 FAPPEFOVL 93  
 QY 1 FAMPNFOVL 9

RESULT 14 PRELIMINARY; PRT; 300 AA.

ID 044173  
 AC 044173;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE PORTION OF HYPOTHETICAL PROTEIN.  
 OS AGNEMELUM OUDRUPICATUM.  
 CC PROKARYOTA; BACTERIA; GRACILICUTES; OXYPHOTOBACTERIA;  
 OC CYANOBACTERIA.  
 RN [1]

SEQUENCE FROM N.A.  
 RC STRAIN-PR-6;  
 RX MEDLINE: 93139025.  
 RA WAGNER S.J., THOMAS S.P., KAUFMAN R., NIXON B., STEVENS S.;  
 RL J. BACTERIOL. 175:604-612(1993).  
 DR EMBL: Z13965; G669044;  
 SQ SEQUENCE 300 AA; 33005 MW; E103F999 CRC32;

Query Match 70.0%; Score 49; DB 9; Length 300;  
 Best Local Similarity 55.6%; Pred. No. 1.33e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 21 FMPKFRNL 29  
 QY 1 FAMPNFOVL 9

RESULT 15 PRELIMINARY; PRT; 223 AA.

ID 060970  
 AC 060970;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE PROTEIN TYROSINE PHOSPHATASE-LIKE.  
 OS MUS MUSCULUS (MOUSE).  
 CC EURAROTIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]

SEQUENCE FROM N.A.  
 RC TISSUE-SKELETAL MUSCLE;  
 RX MEDLINE: 96070766.  
 RA WISHART M.J., DENDU J.M., WILLIAMS J.A., DIXON J.E.;  
 RL J. BIOL. CHEM. 270:26782-26785(1995).  
 DR EMBL: U34973; G1063626;  
 SQ SEQUENCE 223 AA; 25416 MW; 7D7F6D83 CRC32;

Query Match 66.7%; Score 48; DB 10; Length 223;  
 Best Local Similarity 66.7%; Pred. No. 2.06e+01;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 71 FMPNFOVL 79

Sun Sep 13 10:56:58 1998

US-08-452-843-5.rspt

Page 5







OY 1 :|||||  
FAMNFOPL 9

Search completed: Fri Sep 11 12:55:56 1998  
Job time : 39 secs.

---

**THIS PAGE BLANK (USPTO)**



Release 3.1A John F. Collins, Biocomputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

MPSrch_pp  protein - protein database search, using Smith-Waterman algorithm
Run on:      Fri Sep 11 12:49:34 1998; MasPar time 2.61 Seconds
              55,814 Million cell updates/sec
Tabular output not generated.

```

```
Title: >US-08-452-843-4
Description: (1-9) from US08452843.pep
Perfect Score: 80
Sequence: 1 YPAETLLYW 9
```

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

```
post-processing: Minimum Match 08
                  Listing first 45 summaries
```

```
Database:
a-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29
```

Statistics: Mean 18.306; Variance 56.308; scale 0.325

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| No. | Score | Query Match | Length | DB | ID     | Description           | Pred. No |
|-----|-------|-------------|--------|----|--------|-----------------------|----------|
| 1   | 80    | 100.0       | 9      | 18 | R89365 | B53 self peptide det1 | 7.51e-02 |
| 2   | 64    | 80.0        | 9      | 15 | R78845 | B53 self peptide cyto | 4.76e-00 |
| 3   | 64    | 80.0        | 274    | 1  | P80911 | Consensus sequence of | 4.76e-00 |
| 4   | 64    | 80.0        | 337    | 2  | P70590 | Sequence of the human | 4.76e-00 |
| 5   | 64    | 80.0        | 352    | 2  | R03142 | Sequence of HLA-Bw52  | 4.76e-00 |
| 6   | 64    | 80.0        | 352    | 3  | R12463 | HLA-Bw53 exon.        | 4.76e-00 |
| 7   | 64    | 80.0        | 352    | 2  | R03144 | Sequence of HLA-B51 a | 4.76e-00 |
| 8   | 64    | 80.0        | 352    | 2  | P70155 | Sequence encoded by g | 4.76e-00 |
| 9   | 64    | 80.0        | 362    | 3  | R12464 | HLA-B*35 antigen.     | 4.76e-00 |
| 10  | 64    | 80.0        | 366    | 3  | R12465 | HLA-C exon Cp-2.      | 4.76e-00 |
| 11  | 64    | 80.0        | 366    | 3  | R12465 | HLA-C exon Cp-1.      | 4.76e-00 |
| 12  | 63    | 78.8        | 25     | 13 | R63622 | MHC-I peptide Db-(197 | 6.13e-00 |
| 13  | 63    | 78.8        | 25     | 13 | R71423 | Human MHC I alpha 3 d | 6.13e-00 |
| 14  | 63    | 78.8        | 121    | 10 | R52863 | Mouse MHC alpha-3 dom | 2.70e+01 |
| 15  | 57    | 71.3        | 461    | 16 | R75365 | Phyrase.              | DM beta. |
| 16  | 55    | 70.0        | 263    | 15 | R80832 | Membrane type matrix  | 4.39e+01 |
| 17  | 55    | 68.8        | 604    | 24 | M10640 |                       | 5.58e+01 |
| 18  | 54    | 67.5        | 2466   | 13 | R71498 | Human protein tyrosin |          |

|    |    |      |         |        |                       |          |
|----|----|------|---------|--------|-----------------------|----------|
| 19 | 53 | 65.3 | 468.5   | R25597 | PHO.                  | 7.98e+01 |
| 19 | 52 | 65.0 | 238.18  | R93165 | Anti-rhesus D recombi | 8.08e+01 |
| 21 | 52 | 65.0 | 474.3   | R14676 | Rabbit vitronectin-11 | 8.98e+01 |
| 22 | 52 | 65.0 | 760.24  | W22213 | Human transferrin rec | 8.98e+01 |
| 23 | 51 | 63.8 | 96.12   | R65457 | T-cell receptor V-bet | 1.14e+02 |
| 24 | 51 | 63.8 | 667.29  | W46713 | 80 kDa VIP1(a) toxin  | 1.14e+02 |
| 25 | 51 | 63.8 | 667.28  | W19510 | B. cereus 80 kd vip1a | 1.14e+02 |
| 26 | 51 | 63.8 | 667.17  | R91240 | B. cereus VIP1 protel | 1.14e+02 |
| 27 | 51 | 63.8 | 852.12  | R63794 | Bacillus cereus 80 kd | 1.14e+02 |
| 28 | 51 | 63.8 | 852.29  | W46727 | Malte optimsed VIP1A  | 1.14e+02 |
| 29 | 51 | 63.8 | 832.28  | W19516 | Malte optimsed-B. ce  | 1.14e+02 |
| 30 | 51 | 63.8 | 852.17  | R91246 | VIP1(a) protein with  | 1.14e+02 |
| 31 | 51 | 63.8 | 884.28  | W19509 | B. cereus VIP1(a) pr  | 1.14e+02 |
| 32 | 51 | 63.8 | 884.29  | W46712 | 100 kDa VIP1(a) toxi  | 1.14e+02 |
| 33 | 51 | 63.8 | 884.17  | R91239 | B. cereus VIP1(a) in  | 1.14e+02 |
| 34 | 51 | 63.8 | 884.12  | R63793 | Bacillus cereus 100 k | 1.14e+02 |
| 35 | 51 | 63.8 | 1338.29 | W46731 | VIP2(a)/VIP1(a) fus   | 1.14e+02 |
| 36 | 51 | 63.8 | 1338.28 | W19520 | Malte optimsed-B. ce  | 1.14e+02 |
| 37 | 51 | 63.8 | 1338.17 | R91247 | VIP2(a)-VIP1(a) pro   | 1.14e+02 |
| 38 | 51 | 63.8 | 1346.28 | W46573 | VIP1(a)/VIP2(a) fus   | 1.14e+02 |
| 39 | 51 | 63.8 | 1346.28 | W19513 | B. cereus VIP1(a)/VI  | 1.14e+02 |
| 40 | 51 | 63.8 | 1346.17 | R91245 | VIP2(a) and VIP1(a)   | 1.14e+02 |
| 41 | 50 | 62.5 | 13.22   | W10450 | Human growth hormone  | 1.44e+02 |
| 42 | 50 | 62.5 | 13.22   | W10440 | Human growth hormone  | 1.44e+02 |
| 43 | 50 | 62.5 | 624.17  | R77674 | Glucoamylase from Arx | 1.44e+02 |
| 44 | 49 | 61.3 | 287.16  | W27146 | CH3/Fas ligand domain | 1.81e+02 |
| 45 | 49 | 61.3 | 492.29  | W42338 | CD80-19-alpha-tp fusi | 1.81e+02 |

## ALIGNMENTS

|        |  |                          |
|--------|--|--------------------------|
| RESULT | 1  |                          |
| ID     | R89365   | standard; peptide; 9 AA. |
| AC     | R89365   |                          |
| DC     | 18-SEP-1996  | (first entry)            |
| DE     | B53 self peptide derived immunogenic peptide.                        |                          |
| KW     | Immunogenic peptide; supermotif; HLA molecule; CTL response;         |                          |
| KW     | therapeutic; diagnostic; cancer; viral infection; hepatitis B;       |                          |
| OS     | hepatitis C.   |                          |
| PN     | Synthetic.   |                          |
| PN     | MO603140-A1.   |                          |
| PD     | 08-FEB-1996.   |                          |
| PF     | 21-JUL-1995; 009234.   |                          |
| PR     | 21-JUL-1994; US-278634.  |                          |
| PR     | 23-NOV-1994; US-344824.  |                          |
| PR     | 30-MAY-1995; US-452843.  |                          |
| PA     | (CYTE-) CYTEL CORP.  |                          |
| PI     | Settle A; Sidney J;  |                          |
| DR     | WPI; 96-116784/12.   |                          |
| PT     | Compn. comprising immunogenic peptide with supermotif allowing more  |                          |
| PT     | than one HLA mol. to bind - used to induce CTL response in patient   |                          |
| PT     | and for in vivo and ex vivo therapeutic and diagnostic applications  |                          |
| PS     | Claim 2; Page 26; 32pp; English.                                     |                          |
| CC     | The sequences given in R89362-82 are immunogenic peptides which were |                          |
| CC     | used in the composition of the invention. The composition comprises  |                          |
| CC     | an immunogenic peptide of 9-10 residues with a supermotif which      |                          |
| CC     | allows binding of more than one HLA molecule. It pref. comprises     |                          |
| CC     | two conserved residues, a first at the 2nd position from the N-      |                          |
| CC     | terminal is Pto, and a 2nd at the C-terminal is Met. These peptides  |                          |
| CC     | are used to induce a CTL response in a patient. They are also        |                          |
| CC     | useful in compositions for in vivo and ex vivo therapeutic and       |                          |
| CC     | diagnostic applications, e.g the treatment of cancer and viral       |                          |
| CC     | infections, e.g. hepatitis B and C.                                  |                          |
| CC     | Sequence 9 AA;   |                          |

|                       |         |                     |           |               |
|-----------------------|---------|---------------------|-----------|---------------|
| Query Match           | 100.0%; | Score 80;           | DB 18;    | Length 9;     |
| Best Local Similarity | 100.0%; | Pred. No. 7.51e-02; |           |               |
| Matches               | 9;      | Conservative        | 0;        | Mismatches 0; |
|                       |         |                     | Indels 0; | Gaps 0;       |

## RESULT 2

ID R78845 standard; peptide: 9 AA.  
 AC R78845;  
 DT 27-MAR-1996 (first entry)  
 DE B53 self peptide cytotoxic T lymphocyte epitope.  
 KW lymphocyte; viruses; parasites; tumours; antigens; treatment;  
 OS disease prevention.  
 PN Homo sapiens.  
 PD WO9522317-A1.  
 PF 04-AUG-1995.  
 PR 16-FEB-1995; U02121.  
 PA 16-FEB-1994; US-197484.  
 PI (CYTE-1) CYTEL CORP.  
 PI Cent KW, Grey H, Settle AD, Vitello MA;  
 DR WPI: 95-302545/39.  
 PT Compn. inducing cytotoxic T lymphocyte response to pref. viral,  
 PT bacterial, parasitic or tumour antigens - useful in the treatment  
 PT and prevention of diseases associated with the antigen e.g.  
 hepatitis B.  
 PS Disclosure: Page 17; 109pp; English.  
 CC A compn. which induces a cytotoxic T lymphocyte (CTL) response to  
 CC an antigen (Ag) in a mammal comprises, a CTL Ag response inducing  
 CC peptide (i.e. R78824-R78853) and a lipid conjugated helper T cell  
 CC inducing peptide. The compn. induces a CTL response to bacterial,  
 CC viral or tumour Ags, and is therefore useful in the treatment and  
 CC prevention of diseases associated with the Ag.  
 SQ Sequence 9 AA;

## Query Match

Best Local Similarity 88.9%; Score 64; DB 15; Length 9;  
 Pred. No. 4.76e+00;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 ypaetltw 9  
 |||||||

Oy 1 ypaetltw 9

RESULT 3  
 ID P80911 standard; protein: 274 AA.  
 AC P80911;  
 DT 18-SEP-1990 (first entry)

DE Consensus sequence of peptides which constitute the alpha-1, alpha-2 and  
 DE alpha-3 regions of a class I HLA molecule  
 KW HLA-A2 epitopes; extracellular domains alpha-1, alpha-2 and alpha-3.  
 OS Homo sapiens.  
 PN Key  
 PD Location/Qualifiers  
 PF 1..90  
 PT /note="alpha-1 region"

FT region  
 /note="alpha-2 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

FT region  
 /note="alpha-3 region"

Best Local Similarity 88.9%; Score 64; DB 2; Length 362;  
 Pred. No. 4.76e+00;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 209 ypaetltw 217  
 |||||||

Oy 1 ypaetltw 9

RESULT 4  
 ID P70590 standard; protein: 337 AA.  
 AC P70590;  
 DT 10-APR-1991 (first entry)

DE Sequence of the human histocompatibility antigen HLA B27.  
 KW Rheumatic disorder; genetic screening; diagnosis;  
 KW ankylosing spondylitis.  
 OS Homo sapiens.  
 PN DE3542024-A.  
 PD 04-JUN-1987.  
 PF 28-NOV-1985; DE-542024.  
 PR 28-NOV-1985; DE-542024.  
 PR 21-DEC-1985; DE-545576.  
 PA (BEHW) BEHRINGER AG.  
 PI Rietmuller G, Meo T, Weiss E, Szols H;  
 DR WPI: 87-157893/23.  
 DR N-PSDB: N70935.

PT DNA coding for antigen HLA B27 - and diagnostic reagents contg.  
 PT such DNA, antigen or antibody  
 PS Disclosure: Page 5; 5pp; German.  
 CC The DNA may be used as a hybridisation probe for detecting the HLA  
 CC B27 gene, eg for assessing susceptibility to rheumatic disorders  
 CC such as ankylosing spondylitis, or may be used to transform cells  
 CC for prodn. of HLA B27. The HLA B27 may be used to detect HLA B27  
 CC antibody in human serum, or to produce mono- or polyclonal HLA B27  
 CC antibodies for use in immunoassay.  
 SQ Sequence 337 AA;

Query Match  
 Best Local Similarity 88.9%; Score 64; DB 2; Length 337;  
 Pred. No. 4.76e+00;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 208 ypaetltw 216  
 |||||||

Oy 1 ypaetltw 9

RESULT 5  
 ID R03142 standard; protein: 362 AA.  
 AC R03142;  
 DT 19-MAR-1991 (first entry)

DE Sequence of HLA-B\*52 antigen.  
 KW Probe: HLA class I DNA; immunogen.  
 OS Homo sapiens.  
 PN EP-354580-A.  
 PD 14-FEB-1990.  
 PF 10-AUG-1989.  
 PR 11-AUG-1988; JP-200758.  
 PA (OLYU) Olympus Optical Co., Ltd.  
 PI Kano K, Takiguchi;  
 DR WPI: 90-046289/07.

PT New DNA for class I human leucocyte antigens and derived probes and  
 PT transformed cells, useful for DNA typing, as immunogens etc.  
 PS Disclosure: Page 13; 23pp; English.  
 CC The HLA class I DNA can be used as a source of probes for use in DNA  
 CC typing. Transformed cells, which are useful as immunogens, can be  
 CC obtained by introducing these DNAs into eucaryotic cells.  
 SQ Sequence 362 AA;

Query Match  
 Best Local Similarity 88.9%; Score 64; DB 2; Length 362;  
 Pred. No. 4.76e+00;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 ypaetltw 241  
 |||||||

Oy 1 ypaetltw 9

Db 233 ypaetltw 241  
 |||||||

OY 1 YPAEITLW 9

RESULT 5  
ID R12463 standard; protein: 362 AA.  
AC R12463;  
DT 29-AUG-1991 (first entry)  
DE HLA-B\*53 exon.  
KW Human leukocyte antigen; probe: major histocompatibility complex;  
MHC; class I.  
OS Homo sapiens.  
PN J03112487-A.  
PD 14-MAY-1991.  
PE 22-SEP-1989: 247697.  
PR 22-SEP-1989: JP-247697.  
PA (OLXU) OLYMPUS OPTICAL KK.  
DR WPI: 91-182991/25.  
N-PSDB: Q12114.  
PT HLA-B\*53 gene; DNA probe and transformant cells - used for immunisation, identifying specificity of antiserum etc.  
PS Claim 2, page 1, 11pp; Japanese.  
CC Probes comprising part of the sequence encoding the protein can be used to identify Class I genes. The DNA can be expressed for immunisation of animals and prodn. of monoclonal antibodies specific for the HLA-B\*53 antigen. See also J03112485 and J03112486.  
SQ Sequence 362 AA;

Query Match 80.0%; Score 64; DB 3; Length 362;  
Best Local Similarity 88.9%; Pred. No. 4.76e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 233 ypaetitlw 241  
OY 1 YPAEITLW 9

RESULT 7  
ID R03144 standard; protein: 362 AA.  
AC R03144;  
DT 19-MAR-1991 (first entry)  
DE Sequence of HLA-B\*51 antigen.  
KW Probe; HLA class I DNA; immunogen.  
OS Homo sapiens.  
PN EP-354580-A.  
PD 14-FEB-1990.  
PE 10-AUG-1989.  
PR 11-AUG-1988; JP-200758.  
PA (OLXU) OLYMPUS OPTICAL Co., Ltd.  
PI Kano K. Takiguchi;  
DR WPI: 90-046289/07.  
PT New DNA for class I human leucocyte antigens and derived probes and transformed cells, useful for DNA typing, as immunogens etc.  
PS Disclosure; Pages 12-13; 23pp; English.  
CC The HLA class I DNA can be used as a source of probes for use in DNA typing. Transformed cells, which are useful as immunogens, can be obtained by introducing these DNAs into eucaryotic cells.  
SQ Sequence 362 AA;

Query Match 80.0%; Score 64; DB 2; Length 362;  
Best Local Similarity 88.9%; Pred. No. 4.76e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 233 ypaetitlw 241  
OY 1 YPAEITLW 9

RESULT 8  
ID P70155 standard; protein: 362 AA.  
AC P70155;  
DT 10-MAR-1993 (revised)  
DT 03-APR-1991 (first entry)

DE Sequence encoded by genomic DNA encoding human histocompatibility antigen HLA-B 27.  
KW Ankylosing spondylitis; rheumatic disorder; diagnosis.  
OS Homo sapiens.  
PN EP-226069-A.  
PD 24-JUN-1987.  
PE 21-NOV-1986; 116139.  
PR 01-JAN-1985; DE-542024.  
PR 21-DEC-1985; DE-545576.  
PA (BEHW) BEHRINGER AG.  
PI Szots H, Weiss E, Dornier C, Lang M, Meo T, Riethmuller G;  
DR WPI: 87-171469/25.  
N-PSDB: N70225.  
PT DNA coding for human histocompatibility antigen HLA-B 27 - useful for diagnosis and antigen and antibody prodn.  
PS Disclosure; p6; 13pp; German.  
CC The DNA may be used to detect the HLA-B 27 gene (opt. mutated) in human genetic material. The HLA-B 27 may be used to detect anti-HLA-B 27 antibodies in human serum. The antibodies may be used to determine HLA-B 27 levels in human serum, eg for diagnosis of rheumatic disorders, esp. ankylosing spondylitis.  
SQ Sequence 362 AA;

Query Match 80.0%; Score 64; DB 2; Length 362;  
Best Local Similarity 88.9%; Pred. No. 4.76e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 233 ypaetitlw 241  
OY 1 YPAEITLW 9

RESULT 9  
ID R12464 standard; protein: 362 AA.  
AC R12464;  
DT 29-AUG-1991 (first entry)  
DE HLA-B\*35 antigen.  
KW Human leukocyte antigen; probe: major histocompatibility complex;  
MHC; class I.  
OS Homo sapiens.  
PN J03112486-A.  
PD 14-MAY-1991.  
PE 22-SEP-1989; 247697.  
PR 22-SEP-1989; JP-247697.  
PA (OLXU) OLYMPUS OPTICAL KK.  
DR WPI: 91-182991/25.  
N-PSDB: Q12115.  
PT HLA-B\*35 gene - used in DNA probe and transformant cells for immunising animals, for developing monoclonal antibody.  
PS Claim 1; Page 1; 11pp; Japanese.  
CC Probes comprising part of the sequence encoding this sequence can be used to identify Class I genes. The DNA can be expressed for immunisation of animals and prodn. of monoclonal antibodies specific for the HLA-B\*35 antigen. See also J03112485 and J03112487.  
SQ Sequence 362 AA;

Query Match 80.0%; Score 64; DB 3; Length 362;  
Best Local Similarity 88.9%; Pred. No. 4.76e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 233 ypaetitlw 241  
OY 1 YPAEITLW 9

RESULT 10  
ID R12466 standard; protein: 366 AA.  
AC R12466;  
DT 29-AUG-1991 (first entry)  
DE HLA-C exon Cb-2.  
KW Human leukocyte antigen; probe: major histocompatibility complex;  
MHC; class I.  
OS Homo sapiens.

PN J03112485-A.  
 PD 14-MAY-1991.  
 PE 22-SEP-1989; 247695.  
 PR 22-SEP-1989; JP-247695.  
 PA (OLYU ) OLYMPUS OPTICAL KK.  
 DR WPI: 91-182989/25.  
 N-PSDB: 012117.  
 PT HLA-C gene, DNA probe and transformant cells - for immunisation of animals and monoclonal antibody development.  
 PS Claim 4; Page 2; 13pp; Japanese.  
 CC Probes comprising part of the DNA sequence encoding the protein can be used to identify Class I genes. The DNA can be expressed for immunisation of animals and prodn. of monoclonal antibodies specific for the HLA-C antigen. See also R12465 (same patent) and J03112486 and J03112487.  
 CC and J03112487.  
 SQ Sequence 366 AA;

Query Match 80.0%; Score 64; DB 3; Length 366;  
 Best Local Similarity 88.9%; Pred. No. 4.76e+00;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 ypaefltlw 241  
 |||||  
 1 YPAEITLW 9

RESULT 11  
 ID R12465 standard; Protein; 366 AA.

AC R12465;  
 DT 29-AUG-1991 (first entry)  
 DE HLA-C exon Cb-1.  
 KW Human leukocyte antigen; probe; major histocompatibility complex;  
 KM MHC; class I.  
 OS Homo sapiens.  
 PN J03112485-A.  
 PD 14-MAY-1991.  
 PE 22-SEP-1989; 247695.  
 PR 22-SEP-1989; JP-247695.  
 PA (OLYU ) OLYMPUS OPTICAL KK.  
 DR WPI: 91-182989/25.  
 N-PSDB: 012116.  
 PT HLA-C gene, DNA probe and transformant cells - for immunisation of animals and monoclonal antibody development.  
 PS Claim 3; Page 2; 13pp; Japanese.  
 CC Probes comprising part of the DNA sequence encoding the protein can be used to identify Class I genes. The DNA can be expressed for immunisation of animals and prodn. of monoclonal antibodies specific for the HLA-C antigen. See also R12466 (same patent) and J03112486 and J03112487.  
 CC and J03112487.  
 SQ Sequence 366 AA;

Query Match 80.0%; Score 64; DB 3; Length 366;  
 Best Local Similarity 88.9%; Pred. No. 4.76e+00;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 ypaefltlw 241  
 |||||  
 1 YPAEITLW 9

RESULT 12  
 ID R69622 standard; peptide; 25 AA.

AC R69622;  
 DT 29-AUG-1995 (first entry)  
 DE MHC-I peptide Db-(197-221).  
 KW MHC class I; major histocompatibility complex; insulin receptor; diabetes; glucose uptake; adipocyte.  
 OS Synthetic.  
 PN US5385888-A.  
 PD 31-JAN-1995.  
 PE 20-MAR-1987; 028241.  
 PR 20-MAR-1987; US-028241.  
 PR 14-MAR-1989; US-323565.

PR 01-FEB-1991; US-649471.  
 PR 03-MAY-1993; US-057184.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Goodenow RS, Olsson L;  
 DR WPI: 95-081582/11.  
 PT Modulating response of cellular insulin receptor to ligand - using peptide deriv. from MHC class I antigen, partic. to potentiate effect of insulin for treating diabetes  
 PT using peptide deriv. from MHC class I antigen, partic. to potentiate effect of insulin for treating diabetes  
 PS Disclosure; Column 18; 15pp; English.  
 CC Response of an insulin receptor (IR) to a ligand is modulated by contacting mammalian cells having IR on the surface with peptides derived from MHC class I antigen. Peptide Dk-(61-85) (R69619), CC from the alpha-1 domain of MHC-I, boosted glucose uptake 5-6 fold over basal levels in rat adipocytes, when administered at 30 uM.  
 CC Peptide Db-(191-221), from the alpha-3 region, had little or no effect.  
 CC  
 SQ Sequence 25 AA;

Query Match 78.8%; Score 63; DB 13; Length 25;  
 Best Local Similarity 77.8%; Pred. No. 6.13e+00;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 13 ypadltlw 21  
 |||||  
 1 YPAEITLW 9

RESULT 13  
 ID R71423 standard; peptide; 25 AA.

AC R71423;  
 DT 12-OCT-1995 (first entry)  
 DE Human MHC I alpha 3 domain peptide Dk-(197-221).  
 KW Major histocompatibility complex class I; MHC I; EGF receptor; alpha 3 domain; peptide Dk-(197-221); interaction modulation;  
 KM arthritis; neoplasias; lupus erythematosus.  
 OS Homo sapiens.  
 PN W09505189-A.  
 PD 23-FEB-1995.  
 PE 12-AUG-1994; 009189.  
 PR 12-AUG-1993; US-105416.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Goldstein A, Goodenow RS, Olsson L;  
 DR WPI: 95-098577/13.  
 PT Regulating cell surface receptor response - by modulating interaction between MHC class I antigen and the cell surface receptor  
 PT interaction between MHC class I antigen and the cell surface receptor  
 PS Example 3; Page 38; 103pp; English.  
 CC R71420-R71423 are human major histocompatibility complex class I (MHC I) derived peptides, they were used to modulate interactions between MHC I and EGF cell surface receptors. Via competitive inhibition the peptide diminishes the receptors response, this feature may be useful for the treatment of neoplasias, lupus erythematosus and arthritis.  
 CC  
 SQ Sequence 25 AA;

Query Match 78.8%; Score 63; DB 13; Length 25;  
 Best Local Similarity 77.8%; Pred. No. 6.13e+00;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 13 ypadltlw 21  
 |||||  
 1 YPAEITLW 9

RESULT 14  
 ID R52863 standard; Protein; 121 AA.

AC R52863;  
 DT 09-SEP-1994 (first entry)  
 DE Mouse MHC alpha-3 domain.  
 KW MHC; major histocompatibility; target binding polypeptide; antibody engineering; humanized antibody; pPOW; vector; ss. Mus sp.  
 OS  
 PN W09407921-A.

PD 14-APR-1994.  
 PF 24-SEP-1993: AU0491.  
 PR 25-SEP-1992: AU-004973.  
 PA (CSIR) COMMONWEALTH SCI & IND RES ORG.  
 PI Atwell JL, Colman PM, Hudson PJ, Irving RA, Kortt A;  
 PI Lah M, Malbyrl, Power BE;  
 DR WPI: 94-135515/16.  
 DR P-PSDB: 062955.  
 PT New target-binding polypeptide(s) used for diagnosis, etc. -  
 PT having a stable core polypeptide region with at least one  
 PT target-binding region covalently attached, opt. mutated to alter  
 PT specificity, etc.  
 PS Disclosure: Page 39; 67pp; English.  
 CC Mouse MHC alpha-3 domain construct, designed for expression in a  
 CC bacterial secretion vector such as pPCW, has the DNA sequence  
 CC given in Q62955, encoding the protein given in R52863.  
 SQ Sequence 121 AA;

Query Match 78.8%; Score 63; DB 10; Length 121;  
 Best Local Similarity 77.8%; Pred. No. 6.13e+00;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 46 ypaditltw 54  
 ||:||||  
 QY 1 YPAEITLW 9

RESULT 15  
 ID R75365 standard; Protein; 461 AA.  
 AC R75365;  
 DT 25-APR-1996 (first entry)  
 DE Phytase.  
 KM Phytase; phytate; phytic acid; myo-inositol; feed-additive.  
 OS Schwanniomyces occidentalis.  
 PN A09524840-A.  
 PD 25-JAN-1996.  
 PF 04-JUL-1995: 024840.  
 PR 05-JUL-1994: JP-174906.  
 PR 24-FEB-1995: JP-060111.  
 PA (MITSU) MITSUI TOATSU CHEM INC.  
 PI Mochizuki D, Shimada M, Suzuki T, Tawaki S, Tokuda J;  
 DR WPI: 96-097882/11.  
 DR N-PSDB: T15429.  
 PT Phytase from Schwanniomyces occidentalis comprising a single type of  
 PT sub-unit useful to convert phytate into myo-inositol, for use as  
 PT e.g. a feed additive  
 PS Claim 14; Page 33-36; 53pp; English.  
 CC A CDNA clone (T15429) codes for Schwanniomyces occidentalis phytase  
 CC (R75365), an enzyme useful for converting phytate into myo-inositol.  
 CC The phytase consists of a single subunit, facilitating its prodn.  
 CC by genetic engineering. Preferred hosts for recombinant phytase  
 CC and Saccharomyces cerevisiae FERM P-5109. The phytase has  
 CC residual activity of over 90% after heat treatment at 70 deg for 30  
 CC min. Incorporation of phytase into cattle feedstuffs will decrease  
 CC the phosphate content of the cattle excrement, thereby reducing  
 CC pollution.  
 SQ Sequence 461 AA;

Query Match 71.3%; Score 57; DB 16; Length 461;  
 Best Local Similarity 62.5%; Pred. No. 2.70e+01;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 440 ptevitlyw 447  
 ||:||||  
 QY 2 PAEITLW 9

Search completed: Fri Sep 11 12:49:50 1998  
 Job time : 16 secs.

**THIS PAGE BLANK (USPTO)**

(TM)

Distribution rights by Oxford Molecular Ltd

mslcn\_pp protein - pr

Run on: 110 min

tabular output not generated.

Sequence: 1 YPAEITLYW 9

scoring table: PAM 150

SEARCHED: 120441

## Listing

112:1

STATISTICS: Mean 24

and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description           | Pred. No. |
|------------|-------|-------------|--------|----|--------|-----------------------|-----------|
| 1          | 68    | 85.0        | 1226   | 2  | H64479 | magnesium chelata s   | 2.67e-01  |
| 2          | 64    | 80.0        | 270    | 5  | IHLAA  | MHC class I histocomp | 1.19e+00  |
| 3          | 64    | 80.0        | 270    | 5  | 3HLAA  | MHC class I histocomp | 1.19e+00  |
| 4          | 64    | 80.0        | 270    | 5  | 2HLAA  | MHC class I histocomp | 1.19e+00  |
| 5          | 64    | 80.0        | 270    | 5  | IHSBA  | MHC class I histocomp | 1.19e+00  |
| 6          | 64    | 80.0        | 274    | 5  | IAD7A  | hla-a 0201 extracellu | 1.19e+00  |
| 7          | 64    | 80.0        | 275    | 5  | 2CIRA  | Human class I histoco | 1.19e+00  |
| 8          | 64    | 80.0        | 275    | 5  | 2CIRD  | Human class I histoco | 1.19e+00  |
| 9          | 64    | 80.0        | 275    | 5  | IHRHA  | class I histocompatib | 1.19e+00  |
| 10         | 64    | 80.0        | 275    | 5  | IHRJD  | MHC class I histocomp | 1.19e+00  |
| 11         | 64    | 80.0        | 275    | 5  | IHRJA  | MHC class I histocomp | 1.19e+00  |
| 12         | 64    | 80.0        | 275    | 5  | IHRGA  | MHC class I histocomp | 1.19e+00  |
| 13         | 64    | 80.0        | 275    | 5  | IHHID  | MHC class I histocomp | 1.19e+00  |
| 14         | 64    | 80.0        | 275    | 5  | IHHIA  | MHC class I histocomp | 1.19e+00  |
| 15         | 64    | 80.0        | 275    | 5  | IHHGD  | MHC class I histocomp | 1.19e+00  |
| 16         | 64    | 80.0        | 275    | 5  | IHHRD  | MHC class I histocomp | 1.19e+00  |
| 17         | 64    | 80.0        | 275    | 5  | IHRKA  | MHC class I histocomp | 1.19e+00  |
| 18         | 64    | 80.0        | 276    | 5  | IACEA  | b0801 extracellular,  | 1.19e+00  |
| 19         | 64    | 80.0        | 276    | 5  | IACDA  | b0801 extracellular,  | 1.19e+00  |
| 20         | 64    | 80.0        | 276    | 5  | IACJA  | MHC class I histocomp | 1.19e+00  |
| 21         | 64    | 80.0        | 276    | 5  | IACFA  | b0801 extracellular,  | 1.19e+00  |
| 22         | 64    | 80.0        | 276    | 5  | IACBA  | b0801 extracellular,  | 1.19e+00  |
| 23         | 64    | 80.0        | 276    | 5  | IHSAA  | MHC class I histocomp | 1.19e+00  |

|    |      |      |     |   |         |                           |          |
|----|------|------|-----|---|---------|---------------------------|----------|
| 25 | 64   | 80.0 | 276 | 5 | HSAD    | MHC class I histocomp     | 1.19e+00 |
| 24 | 64   | 80.0 | 276 | 5 | IAGCA   | OB081 extracellular,      | 1.19e+00 |
| 26 | 64   | 80.0 | 308 | 2 | I36956  | MHC Ch1a chain - chim     | 1.19e+00 |
| 64 | 80.0 |      | 313 | 2 | I36958  | MHC Ch1a chain - chim     | 1.19e+00 |
| 64 | 80.0 |      | 332 | 2 | S064424 | MHC class I histocompatib | 1.19e+00 |
| 28 | 64   | 80.0 | 345 | 2 | S071114 | MHC class I histocomp     | 1.19e+00 |
| 29 | 64   | 80.0 | 354 | 2 | I80170  | MHC class I histocomp     | 1.19e+00 |
| 30 | 64   | 80.0 | 357 | 2 | I11136  | class I histocompatib     | 1.19e+00 |
| 31 | 64   | 80.0 | 357 | 2 | S11134  | class I histocompatib     | 1.19e+00 |
| 32 | 64   | 80.0 | 357 | 2 | S11134  | class I histocompatib     | 1.19e+00 |
| 33 | 64   | 80.0 | 362 | 2 | I84486  | transmembrane glycoprt    | 1.19e+00 |
| 34 | 64   | 80.0 | 362 | 2 | I38421  | gene HLA-B-1519 prote     | 1.19e+00 |
| 35 | 64   | 80.0 | 362 | 2 | I38437  | MHC class I histocomp     | 1.19e+00 |
| 36 | 64   | 80.0 | 362 | 2 | I37521  | HLA-Bw57.2 antigen -      | 1.19e+00 |
| 37 | 64   | 80.0 | 362 | 2 | I56133  | MHC class I protein -     | 1.19e+00 |
| 38 | 64   | 80.0 | 362 | 2 | JH0288  | class I histocompatib     | 1.19e+00 |
| 39 | 64   | 80.0 | 362 | 2 | I37120  | MHC class I histocomp     | 1.19e+00 |
| 40 | 64   | 80.0 | 362 | 1 | HLHDB8  | MHC class I histocomp     | 1.19e+00 |
| 41 | 64   | 80.0 | 362 | 2 | I61864  | MHC HLA-Bw41 chain -      | 1.19e+00 |
| 42 | 64   | 80.0 | 364 | 2 | I72217  | MHC class IB protein      | 1.19e+00 |
| 43 | 64   | 80.0 | 365 | 2 | I38518  | HLA-A-0102 allele - h     | 1.19e+00 |
| 44 | 64   | 80.0 | 365 | 2 | I36961  | MHC class I protein -     | 1.19e+00 |
| 45 | 64   | 80.0 | 381 | 2 | S35940  | class I histocompatib     | 1.19e+00 |

## ALIGNMENTS

| ENTRY<br>TITLE  | RESULT             | 1  |
|---|--------------------|--|
| H64479<br>magnesium cheilataase subunit homolog - Methanococcus<br>jannaschii | #type complete     |  |
| ORGANISM<br>DATE  | #formal_name       | Methanococcus jannaschii                                   |
| 13-Sep-1996   | #sequence_revision | 13-Sep-1996  |
| 10-Oct-1997   | #text_change       |  |
| A64479  |                    |  |
| A64300  |                    |  |
| ACCESSIONS<br>REFERENCE   |                    |  |
| #authors  |                    | Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, |

```
#journal      Science (1996) 273:1058-1073
#title        Complete genome sequence of the methanogenic archaeon,
               Methanococcus jannaschii.
#cross-references MIMD:96337999
```

```
##status      preliminary; nucleic acid sequence not shown;
               translation not shown
```

```
##residues      1-1226  ##label BUT.
##cross-references GB:U67585; GB:L77117; NID:q1592088; PID:q1500323;
```

TIGR:MJ1441

```
#map_position FOR1408283-1411963
SUMMARY #length 1226 #molecular-weight 141327 #checksum 5085
```

|                       |        |                     |       |               |
|-----------------------|--------|---------------------|-------|---------------|
| Query Match           | 85.08; | Score 68;           | DB 2; | Length 1226;  |
| Best Local Similarity | 66.78; | Pred. NO. 2.67e-01; |       |               |
| Matches               | 6;     | Conservative        | 2;    | Mismatches 1; |
|                       |        |                     |       | Indels 0;     |
|                       |        |                     |       | Gaps 0;       |

```

Db      845 YPENIALYIW 853
      ||:|:|
QY      1 YPAETITLYW 9

```

| RESULT | 2   |
|--------|---|
| ENTRY  | HLA   |
| TITLE  | MHC class I histocompatibility antigen HLA-A2 alpha chain<br>(with beta-2-microglobulin), chain A - human |

| ORGANISM                                   | #note                               | #authors   | #submision  | #cross-references                   | REFERENCE   | #journal  | #title   | REFERENCE  | #authors  | #journal   | #title  | REFERENCE   | #authors                                      | #journal   | #title   | REFERENCE   | #authors                         | #journal  | #title           | REFERENCE                                     | #authors                     | #journal       | #title   |        |
|--|-------------------------------------|--|---|-------------------------------------|---|---|--|--|---|--|---|---|---|--|--|---|----------------------------------|---|------------------|---|------------------------------|----------------|----------|--------|
| ORGANISM                                   | #note                               | #authors   | #submision  | #cross-references                   | REFERENCE   | #journal  | #title   | REFERENCE  | #authors  | #journal   | #title  | REFERENCE   | #authors                                      | #journal   | #title   | REFERENCE   | #authors                         | #journal  | #title           | REFERENCE                                     | #authors                     | #journal       | #title   |        |
| 1formal_name Homo sapiens #common_name man | 1lymphoblastoid cell line JY A50156 | Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C.                  | submitted to the Brookhaven Protein Data Bank, October 1987 | #cross-references PDB:1HLA TN004189 | Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C. | Nature (1987) 329:506                                       | Structure of the human class I histocompatibility antigen, HLA-A2. | TN004190   | Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C. | Nature (1987) 329:512  | The foreign antigen binding site and t cell recognition regions of class I histocompatibility antigens. | TN004191  | Bjorkman, P.J.; Strominger, J.L.; Wiley, D.C. | J. Mol. Biol. (1985) 186:205                                       | Crystallization and x-ray diffraction studies on the histocompatibility antigens HLA-A2 and HLA-A28 from human cell membranes. | Resolution: 3.5 angstroms   | Determination: X-ray diffraction | histocompatibility antigen  | #disulfide bonds | #length 270                                   | #molecular-weight 31122      | #checksum 6344 |          |        |
| Query Match                                | Best Local Similarity 88.9%;        | Matches 8;   | Conservative 0;   | Mismatches 1;                       | Indels 0;   | Gaps 0  |  |  |   |  |   |   |   |  |  |   |                                  |   |                  |   |                              |                |          |        |
| Db 209                                     | YP81TLTW 217                        |  | 1   | YP81TLTW 9                          |   |   |  |  |   |  |   |   |   |  |  |   |                                  |   |                  |   |                              |                |          |        |
| RESULT 3                                   | ENTRY                               | TITLE  | ORGANISM  | #note                               | REFERENCE   | #authors  | #submision   | #cross-references  | REFERENCE   | #authors   | #journal  | #title  | REFERENCE                                     | #authors   | #journal   | #title  | REFERENCE                        | #authors  | #journal         | #title  | REFERENCE                    | #authors       | #journal | #title |
| 3HLA                                       | #type complete                      | MHC class I histocompatibility antigen HLA-A2.1 alpha chain (with beta-2-microglobulin), chain A - human | 1formal_name Homo sapiens #common_name man                  | 1lymphoblastoid cell line JY A50601 | Saper, M.A.; Bjorkman, P.J.; Wiley, D.C.  | submitted to the Brookhaven Protein Data Bank, October 1989 | #cross-references PDB:3HLA TN004711                                | Garett, T.P.J.; Saper, M.A.; Bjorkman, P.J.; Strominger, J.L.; Wiley, D.C. | Nature (1989) 342:692   | Specificity pockets for the side chains of peptide antigens in HLA-Aw68. | TN004712  | Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C. | Nature (1987) 329:506                         | Structure of the human class I histocompatibility antigen, HLA-A2. | TN004713   | Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C. | Nature (1987) 329:512            | The foreign antigen binding site and t cell recognition regions of class I histocompatibility antigens. | TN004714         | Bjorkman, P.J.; Strominger, J.L.; Wiley, D.C. | J. Mol. Biol. (1985) 186:205 |                |          |        |

|                       |  |
|-----------------------|--|
| #title                | Crystallization and x-ray diffraction studies on the histocompatibility antigens HLA-A2 and HLA-A28 from human cell membranes. |
| COMMENT               | Resolution: 2.6 angstroms  |
| COMMENT               | R-value: 0.169   |
| COMMENT               | Determination: X-ray diffraction   |
| FEATURE               |  |
| 46-47, 31-37, 21-28,  |  |
| 3-12, 94-103,         |  |
| 109-118, 121-126,     |  |
| 133-135               | #region beta sheet\  |
| 50-53                 | #region helix (right hand alpha) (h1 alpha 1 in jrn1)\   |
| 57-84                 | #region helix (right hand alpha) (h2 alpha 1 in jrn1)\   |
| 138-150               | #region helix (right hand alpha) (h1 alpha 2 in jrn1)\   |
| 132-161               | #region helix (right hand alpha) (h2 alpha 2 in jrn1)\   |
| 153-174               | #region helix (right hand alpha) (h2b alpha 2 in jrn1)\  |
| 176-179               | #region helix (right hand alpha) (h3 alpha 2 in jrn1)\   |
| 186-193, 198-208,     |  |
| 241-250, 228-230      | #region beta sheet\  |
| 186-193, 198-208,     |  |
| 241-250, 234-235      | #region beta sheet\  |
| 222-224, 214-219,     |  |
| 257-262               | #region beta sheet\  |
| 101-164               | #disulfide_bonds\  |
| 203-259               | #disulfide_bonds\  |
| SUMMARY               | #length 270 #molecular-weight 31122 #checksum 6344   |
| Query Match           | 80.0%; Score 64; DB 5; Length 270;   |
| Best Local Similarity | 88.9%; Pred. No. 1.19e+00;   |
| Matches               | 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0  |
| Db                    | 209 YPAAITLW 217   |
| QY                    | 1 YPAAITLW 9   |
| RESULT                | 4  |
| ENTRY                 | 2HLAA  |
| TITLE                 | MHC class I histocompatibility antigen HLA-AW 68.1 alpha chain (with beta-2-microglobulin), chain A - human                    |
| ORGANISM              | #formal_name Homo sapiens #common_name human   |
| note                  | lymphoblastoid cell line LB  |
| REFERENCE             | A50458   |
| authors               | Garrett, T.P.J.; Saper, M.A.; Wiley, D.C.  |
| #submission           | submitted to the Brookhaven Protein Data Bank, October 1989  |
| #cross-references     | PDB:2HLA   |
| REFERENCE             | TN017835   |
| authors               | Garrett, T.P.J.; Saper, M.A.; Bjorkman, P.J.; Strominger, J.L.; Wiley, D.C.  |
| #journal              | Nature (1989) 342:692  |
| #title                | Specificity pockets for the side chains of peptide antigens in HLA-Aw68.   |
| REFERENCE             | TN017836   |
| authors               | Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C.  |
| #journal              | Nature (1987) 329:506  |
| #title                | Structure of the human class I histocompatibility antigen, HLA-A2.   |
| REFERENCE             | TN017837   |
| authors               | Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C.  |
| #journal              | Nature (1987) 329:512  |
| #title                | The foreign antigen binding site and t cell recognition regions of class I histocompatibility antigens.                        |
| REFERENCE             | TN017838   |
| authors               | Bjorkman, P.J.; Strominger, J.L.; Wiley, D.C.  |
| #journal              | J. Mol. Biol. (1985) 186:205   |
| #title                | Crystallization and x-ray diffraction studies on the histocompatibility antigens HLA-A2 and HLA-A28 from human cell membranes. |
| COMMENT               | Resolution: 2.6 angstroms  |
| COMMENT               | R-value: 0.173   |
| COMMENT               | Determination: X-ray diffraction   |



```

FEATURE
28-31      #region turn\
46-47,30-37,21-28,
3-12,93-103,
109-118,121-126,
133-135      #region beta sheet\
49-52      #region turn\
52-55      #region turn\
56-59      #region turn\
57-84      #region helix (right hand alpha)\
83-86      #region turn\
118-121    #region turn\
127-130    #region turn\
138-149    #region helix (right hand alpha)\
151-154    #region turn\
152-161    #region helix (right hand alpha)\
163-175    #region helix (right hand alpha)\
173-176    #region turn\
178-181    #region turn\
183-183,198-209,
241-250,228-230,
186-193,198-209,
241-250,234-235,
219-222    #region beta sheet\
222-224,213-219,
257-263    #region beta sheet\
225-228    #region turn\
250-253    #region turn\
263-266    #region turn\
101-164    #disulfide_bonds\
203-259    #disulfide_bonds\

SUMMARY      #length 270 #molecular-weight 31109 #checksum 6553

Query Match      80.0%; Score 64; DB 5; Length 270;
Best Local Similarity 88.9%; Pred. No. 1.19e+00;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 209 YPAEITLW 217
QY 1 YPAEITLW 9

RESULT 5
ENTRY      1HSBA      #type complete
TITLE      MHC class I histocompatibility antigen HLA-A*68.1 alpha
            chain (leucocyte antigen), chain A - human
ORGANISM    #formal_name Homo sapiens #common_name man
            #lymphoblastoid cell line lb
REFERENCE    A51236
            Guo, H.C.; Strominger, J.L.; Wiley, D.C.
            #authors
            #submitted to the Brookhaven Protein Data Bank, March 1993
            #cross-references PDB:1HSB
REFERENCE    TN016675
            Guo, H.C.; Jardeitzky, T.S.; Garrett, T.P.J.; Lane, W.S.;
            Strominger, J.L.; Wiley, D.C.
            #journal
            #title Nature (1992) 360:364
            Different length peptides bind to HLA-A*68 similarly at their
            ends but bulge out in the middle.
            TN016676
            Guo, H.C.; Jardeitzky, T.S.; Garrett, T.P.J.; Lane, W.S.;
            Strominger, J.L.; Wiley, D.C.
            #journal
            #title Nature (1992) 360:367
            Atomic structure of a human mhc molecule presenting an
            influenza virus peptide.
            A43338
            Madden, D.R.; Gorga, J.C.; Strominger, J.L.; Wiley, D.C.
            #authors
            #journal Cell (1992) 70:1035
            The three-dimensional structure of hla-b27 at 2.1 angstroms

```

```

resolution suggests a general mechanism for tight peptide
binding to mhc.
#cross-references M0ID:92405152
REFERENCE    A58495
            Saper, M.A.; Bjorkman, P.J.; Wiley, D.C.
            #authors
            #journal J. Mol. Biol. (1991) 219:277
            #title Refined structure of the human histocompatibility antigen
            hla-a2 at 2.6 angstroms resolution.
            #cross-references M0ID:91245570
REFERENCE    TN016680
            Madden, D.R.; Gorga, J.C.; Strominger, J.L.; Wiley, D.C.
            #authors
            #journal Nature (1991) 353:321
            #title The structure of hla-b27 reveals nonamer self-peptides bound
            in an extended conformation.
REFERENCE    TN016681
            Garrett, T.P.J.; Saper, M.A.; Bjorkman, P.J.; Strominger,
            J.L.; Wiley, D.C.
            #journal Nature (1989) 342:692
            #title Specificity pockets for the side chains of peptide antigens
            in hla-a*68.
REFERENCE    TN016682
            Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.;
            Strominger, J.L.; Wiley, D.C.
            #journal Nature (1987) 329:506
            #title Structure of the human class I histocompatibility antigen,
            HLA-A2.
REFERENCE    TN016683
            Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.;
            Strominger, J.L.; Wiley, D.C.
            #journal Nature (1987) 329:512
            #title The foreign antigen binding site and t cell recognition
            regions of class I histocompatibility antigens.
REFERENCE    TN016684
            Bjorkman, P.J.; Strominger, J.L.; Wiley, D.C.
            #authors
            #journal J. Mol. Biol. (1985) 186:205
            #title Crystallization and x-Ray diffraction studies on the
            histocompatibility antigens HLA-a2 and hla-a28 from human
            cell membranes.
COMMENT      Resolution: 1.9 angstroms
COMMENT      Determination: X-ray diffraction
FEATURE
46-47,31-37,21-28,
3-12,94-103,
109-118,121-126,
133-135      #region beta sheet\
50-53      #region helix (right hand alpha)\
57-84      #region helix (right hand alpha)\
138-150    #region helix (right hand alpha)\
152-161    #region helix (right hand alpha)\
163-174    #region helix (right hand alpha)\
176-179    #region helix (right hand alpha)\
186-193,198-208,
241-250,228-230,
186-193,198-208,
241-250,234-235,
222-224,214-219,
257-262    #region beta sheet\
101-164    #region beta sheet\
203-259    #region beta sheet\
            #region beta sheet\
            #region beta sheet\
            #disulfide_bonds\
            #disulfide_bonds\

SUMMARY      #length 270 #molecular-weight 31109 #checksum 6553

Query Match      80.0%; Score 64; DB 5; Length 270;
Best Local Similarity 88.9%; Pred. No. 1.19e+00;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 209 YPAEITLW 217
QY 1 YPAEITLW 9

RESULT 6
ENTRY      1A07A      #type complete
TITLE      hla-a 0201 extracellular domains alpha 1, alpha 2, alpha 3

```

ALTERNATE\_NAMES beta-2 microglobulin b, chain A - human  
PDB\_TITLE complex between human t-cell receptor, viral peptide (tax)  
#and hla-a 0201

ORGANISM #formal\_name Homo sapiens #common\_name man  
#note expressed in Escherichia coli, strain xag0 inclusion bodies  
#REFERENCE A68017

#authors Garboczi, D.N.; Ghosh, P.; Utz, U.; Fan, Q.R.; Biddison, W.E.; Wiley, D.C.

#submission submitted to the Brookhaven Protein Data Bank, July 1997  
#cross-references PDB:1A07

#REFERENCE TN000413

#journal Garboczi, D.N.; Ghosh, P.; Utz, U.; Fan, Q.R.; Biddison, W.E.; Wiley, D.C.  
#title Nature (1996) 384:134

#journal Structure of the complex between human t-cell receptor, viral peptide and hla-a2.

#authors Garboczi, D.N.; Utz, U.; Ghosh, P.; Seth, A.; Kim, J.; Vantienhoven, E.A.; Biddison, W.E.; Wiley, D.C.  
#journal J. Immunol. (1996) 157:5403

#title Assembly, specific binding, and crystallization of a human tcr-alpha-beta with an antigenic tax peptide from human t lymphotropic virus type 1 and the class I mhc molecule hla-a2.

COMMENT Resolution: 2.6 angstroms  
Determination: X-ray diffraction  
R-value: 0.245  
class I mhc: complex; mhc: receptor; t-cell receptor; viral peptide; viral peptide complex

KEYWORDS

FEATURE

50-55 #region helix (right hand 3-10)\  
57-84 #region helix (right hand alpha)\  
138-149 #region helix (right hand alpha)\  
152-161 #region helix (right hand alpha)\  
163-174 #region helix (right hand alpha)\  
176-179 #region helix (right hand alpha)\  
224-227 #region helix (right hand 3-10)\  
253-256 #region helix (right hand 3-10)\  
32-37,21-27,3-12, #region beta sheet\  
94-103,109-118, #region beta sheet\  
121-126,133-135 #region beta sheet\  
243-249,198-205, #region beta sheet\  
188-195 #region beta sheet\  
214-218,258-262, #region beta sheet\  
270-273

SUMMARY #length 274 #molecular-weight 31679 #checksum 992

Query Match 80.0%; Score 64; DB 5; Length 274;  
Best Local Similarity 88.9%; Pred. No. 1.19e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 209 YPAEITLW 217  
OY 1 YPAEITLW 9

RESULT 7 2CLRA #type complete  
ENTRY Human class I histocompatibility antigen (hla-a 0201)  
TITLE A - synthetic

ORGANISM #formal\_name synthetic  
#note hla-a 0201: human (homo sapiens) recombinant extracellular fragment expressed in Escherichia coli; peptide: synthetic based on the sequence of human calreticulin

REFERENCE A67214  
#authors Collins, E.J.; Garboczi, D.N.; Wiley, D.C.  
#submission submitted to the Brookhaven Protein Data Bank, August 1994  
#cross-references PDB:2CLR TN023775

#authors Collins, E.J.; Garboczi, D.N.; Wiley, D.C.  
#journal Nature (1994) 371:626

#title Three dimensional structure of a peptide extending out one end of a class I mhc binding site.

REFERENCE TN023776  
#authors Garboczi, D.N.; Madden, D.R.; Wiley, D.C.  
#journal J. Mol. Biol. (1994) 239:581

#title Five viral peptide-hla-a2 co-crystals: simultaneous space group determination and x-ray data collection.

REFERENCE TN023777  
#authors Guo, H.C.; Madden, D.R.; Silver, M.L.; Jardelezky, T.S.; Gorga, J.C.; Strominger, J.L.; Wiley, D.C.  
#journal Proc. Natl. Acad. Sci. U.S.A. (1993) 90:8053

#title Comparison of the p2 specificity pocket in three human histocompatibility antigens: hla-a\*6801, hla-a\*0201, and hla-b\*2705.

REFERENCE A49433  
#authors Madden, D.R.; Garboczi, D.N.; Wiley, D.C.  
#journal Cell (1993) 75:693

#title The antigenic identity of peptide-mhc complexes a comparison of the conformations of five viral peptides presented by hla-a2.

REFERENCE TN023779  
#authors Garboczi, D.N.; Hung, D.T.; Wiley, D.C.  
#journal Proc. Natl. Acad. Sci. U.S.A. (1992) 89:3429

#title Hla-a2-peptide complexes: refolding and crystallization of molecules expressed in escherichia coli and complexed with single antigenic peptides.

REFERENCE A58495  
#authors Saper, M.A.; Bjorkman, P.J.; Wiley, D.C.  
#journal J. Mol. Biol. (1991) 219:277

#title Refined structure of the human histocompatibility antigen hla-a2 at 2.6 angstroms resolution.

REFERENCE TN023781  
#authors Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C.  
#journal Nature (1987) 329:506

#title Structure of the human class I histocompatibility antigen, hla-a2.

REFERENCE TN023782  
#authors Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C.  
#journal Nature (1987) 329:512

#title The foreign antigen binding site and t cell recognition regions of class I histocompatibility antigens.

REFERENCE TN023783  
#authors Bjorkman, P.J.; Strominger, J.L.; Wiley, D.C.  
#journal J. Mol. Biol. (1985) 186:205

#title Crystallization and x-ray diffraction studies on the histocompatibility antigens hla-a2 and hla-a28 from human cell membranes.

COMMENT Resolution: 2.0 angstroms  
Determination: X-ray diffraction  
Histocompatibility antigen

FEATURE

49-53 #region helix (right hand alpha)\  
57-84 #region helix (right hand alpha)\  
140-149 #region helix (right hand alpha)\  
152-161 #region helix (right hand alpha)\  
163-174 #region helix (right hand alpha)\  
176-179 #region helix (right hand alpha)\  
46-47,31-37,21-28, #region helix (right hand alpha)\  
3-12,94-103,  
109-118,121-126,  
133-135 #region beta sheet\  
186-195,198-208, #region beta sheet\  
241-250,229-230, #region beta sheet\  
186-195,198-208, #region beta sheet\  
241-250,234-235, #region beta sheet\  
222-223,214-219, #region beta sheet\  
257-262,270-273, #region beta sheet\  
101-164 #disulfide bonds\  
203-259 #disulfide bonds

SUMMARY #length 275 #molecular-weight 31808 #checksum 4235

| Query Match           | 80.0%   | Score 64           | DB 5 | Length 275                      |
|-----------------------|---|--------------------|------|---------------------------------|
| Best Local Similarity | 88.9%   | Pred. NO. 1.19e+00 |      |                                 |
| Matches               | 8   | Conservative       | 0    | Mismatches 1; Indels 0; Gaps 0; |
| Db                    | 209   | YPAEITLW 217       |      |                                 |
| Qy                    | 1   | YPAEITLW 9         |      |                                 |
| RESULT                | 8   |                    |      |                                 |
| ENTRY                 | 2CLR  | #type complete     |      |                                 |
| TITLE                 | Human class I histocompatibility antigen (hla-a 0201)   |                    |      |                                 |
|                       | D complexed with a decameric peptide from calreticulin, chain   |                    |      |                                 |
| ORGANISM              | #formal_name synthetic  |                    |      |                                 |
| #note                 | hla-a 0201: human (homo sapiens) recombinant extracellular  |                    |      |                                 |
|                       | fragment expressed in <i>Escherichia coli</i> ; peptide: synthetic  |                    |      |                                 |
|                       | based on the sequence of human calreticulin   |                    |      |                                 |
| REFERENCE             | A67214  |                    |      |                                 |
| #authors              | Collins, E.J.; Garboczi, D.N.; Wiley, D.C.  |                    |      |                                 |
| #submission           | submitted to the Brookhaven Protein Data Bank, August 1994  |                    |      |                                 |
| #cross-references     | PDB:2CLR  |                    |      |                                 |
| REFERENCE             | TN023802  |                    |      |                                 |
| #authors              | Collins, E.J.; Garboczi, D.N.; Wiley, D.C.  |                    |      |                                 |
| #journal              | Nature (1994) 371:626   |                    |      |                                 |
| #title                | Three dimensional structure of a peptide extending out one end of a class I mhc binding site.   |                    |      |                                 |
| REFERENCE             | TN023803  |                    |      |                                 |
| #authors              | Garboczi, D.N.; Madden, D.R.; Wiley, D.C.   |                    |      |                                 |
| #journal              | J. Mol. Biol. (1994) 239:581  |                    |      |                                 |
| #title                | Five viral peptide-hla-a2 co-crystals: simultaneous space group determination and x-ray data collection.  |                    |      |                                 |
| REFERENCE             | TN023804  |                    |      |                                 |
| #authors              | Guo, H.C.; Madden, D.R.; Silver, M.L.; Tardetzky, T.S.;   |                    |      |                                 |
| #journal              | Gorga, J.C.; Strominger, J.L.; Wiley, D.C.  |                    |      |                                 |
| #title                | Proc. Natl. Acad. Sci. U.S.A. (1993) 90:8053  |                    |      |                                 |
|                       | Comparison of the p2 specificity pocket in three human  |                    |      |                                 |
|                       | histocompatibility antigens: hla-a*6801, hla-a*0201, and  |                    |      |                                 |
|                       | hla-b*2705.   |                    |      |                                 |
| REFERENCE             | A49433  |                    |      |                                 |
| #authors              | Madden, D.R.; Garboczi, D.N.; Wiley, D.C.   |                    |      |                                 |
| #journal              | Cell (1993) 75:693  |                    |      |                                 |
| #title                | The antigenic identity of peptide-mhc complexes a comparison of the conformations of five viral peptides presented by hla-a2.                           |                    |      |                                 |
| REFERENCE             | TN023806  |                    |      |                                 |
| #authors              | Garboczi, D.N.; Hung, D.T.; Wiley, D.C.   |                    |      |                                 |
| #journal              | Proc. Natl. Acad. Sci. U.S.A. (1992) 89:3429  |                    |      |                                 |
| #title                | Hla-a2-peptide complexes: refolding and crystallization of molecules expressed in <i>escherichia coli</i> and complexed with single antigenic peptides. |                    |      |                                 |
| REFERENCE             | A58495  |                    |      |                                 |
| #authors              | Saper, M.A.; Bjorkman, P.J.; Wiley, D.C.  |                    |      |                                 |
| #journal              | J. Mol. Biol. (1991) 219:277  |                    |      |                                 |
| #title                | Refined structure of the human histocompatibility antigen hla-a2 at 2.6 angstroms resolution.   |                    |      |                                 |
| REFERENCE             | TN023808  |                    |      |                                 |
| #authors              | Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.;   |                    |      |                                 |
| #journal              | Strominger, J.L.; Wiley, D.C.   |                    |      |                                 |
| #title                | Nature (1987) 329:506   |                    |      |                                 |
|                       | Structure of the human class I histocompatibility antigen, hla-a2.  |                    |      |                                 |
| REFERENCE             | TN023809  |                    |      |                                 |
| #authors              | Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.;   |                    |      |                                 |
| #journal              | Strominger, J.L.; Wiley, D.C.   |                    |      |                                 |
| #title                | Nature (1987) 329:512   |                    |      |                                 |
|                       | The foreign antigen binding site and t cell recognition regions of class I histocompatibility antigens.   |                    |      |                                 |
| REFERENCE             | TN023810  |                    |      |                                 |
| #authors              | Bjorkman, P.J.; Strominger, J.L.; Wiley, D.C.   |                    |      |                                 |
| #journal              | J. Mol. Biol. (1985) 186:205  |                    |      |                                 |

```
#title      crystallization and x-ray diffraction studies on the histocompatibility antigens hla-a2 and hla-a28 from human cell membranes.
```

```
COMMENT      Resolution: 2.0 angstroms
```

```
COMMENT      Determination: X-ray diffraction
```

```
KEYWORDS     Histocompatibility antigen
```

```
FEATURES
```

```
49-53        #region helix (right hand alpha)\
```

```
57-84        #region helix (right hand alpha)\
```

```
140-149      #region helix (right hand alpha)\
```

```
152-161      #region helix (right hand alpha)\
```

```
163-174      #region helix (right hand alpha)\
```

```
176-179      #region helix (right hand alpha)\
```

```
46-47,31-37,21-28,
```

```
3-12,94-103,
```

```
109-118,121-126,
```

```
133-135      #region beta sheet\
```

```
186-195,198-208,
```

```
241-250,229-230
```

```
186-195,198-208,
```

```
241-250,234-235
```

```
222-223,214-219,
```

```
257-262,270-273
```

```
101-164      #region beta sheet\
```

```
203-259      #disulfide_bonds\
```

```
SUMMARY      #length 275 #molecular_weight 31808 #checksum 4235
```

```
Query Match  80.0%; Score 64; DB 5; Length 275;
```

```
Best Local Similarity 88.9%; Pred. No. 1.19e+00;
```

```
Matches      8; Conservative 0; Mismatches 1; Indels 0; Gaps 0.
```

```
Db           209 YPAAITLW 217
```

```
Oy           1 YPAEITLW 9
```

```
RESULT       9
```

```
ENTRY        class I histocompatibility antigen HLA-A*0201 (with hepatitis B nucleocapsid protein residues 18-27), chain A - human
```

```
TITLE        #formal_name Homo sapiens #common_name man
```

```
ORGANISM      recombinant form expressed in Escherichia coli
```

```
NOTE          A51224
```

```
REFERENCE     Madden, D.R.; Garboczi, D.N.; Wiley, D.C.
```

```
#authors      submitted to the Brookhaven Protein Data Bank, June 1993
```

```
#submission   TN013775
```

```
#cross-references PDB:1HHH
```

```
REFERENCE     TN013775
```

```
#authors      Garboczi, D.N.; Hung, D.T.; Wiley, D.C.
```

```
#journal      Proc. Natl. Acad. Sci. U.S.A. (1992) 89:3429
```

```
#title        HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in Escherichia coli and complexed with T-cell antigenic peptides.
```

```
TN013776
```

```
REFERENCE     Bertoldetti, A.; Chisari, F.V.; Penna, A.; Guilihot, S.; Galati, L.; Missale, G.; Fowler, P.; Schlicht, H.J.; Vitellio, A.; Chesnut, R.C.; Flaccadori, F.; Ferrari, C.
```

```
#journal      J. Virol. (1993) 67:2376
```

```
#title        Definition of a minimal optimal cytotoxic T-cell epitope within the hepatitis B virus nucleocapsid protein.
```

```
A43338
```

```
REFERENCE     Madden, D.R.; Gorga, J.C.; Strominger, J.L.; Wiley, D.C.
```

```
#journal      Cell (1992) 70:1035
```

```
#title        The three-dimensional structure of HLA-B*27 at 2.1 angstroms resolution suggests a general mechanism for tight peptide binding to MHC.
```

```
#cross-references MIMD:92405152
```

```
REFERENCE     AS8495
```

```
#authors      Saper, M.A.; Bjorkman, P.J.; Wiley, D.C.
```

```
#journal      J. Mol. Biol. (1991) 219:277
```

```
#title        Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 angstroms resolution.
```

```
#cross-references MIMD:91245570
```

```
TN013779
```

```
REFERENCE
```

```

REFERENCE      A4338
#authors      Madden, D.R.; Gorga, J.C.; Strominger, J.L.; Wiley, D.C.
#journal      Cell (1992) 70:1035-1048
#title        The three-dimensional structure of HLA-B27 at 2.1 angstrom
              resolution suggests a general mechanism for tight peptide
              binding to MHC.
#cross-references MTID:92405152
REFERENCE      A58495
#authors      Saper, M.A.; Bjorkman, P.J.; Wiley, D.C.
#journal      J. Mol. Biol. (1991) 219:277
#title        Refined structure of the human histocompatibility antigen
              HLA-A2 at 2.6 angstroms resolution.
#cross-references MTID:91245570
REFERENCE      TNO19834
#authors      Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.;
              Strominger, J.L.; Wiley, D.C.
#journal      Nature (1987) 329:506
#title        Structure of the human class I histocompatibility antigen,
              HLA-A2.
REFERENCE      TNO19835
#authors      Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.;
              Strominger, J.L.; Wiley, D.C.
#journal      Nature (1987) 329:512
#title        The foreign antigen binding site and T cell recognition
              regions of class I histocompatibility antigens.
REFERENCE      TNO19836
#authors      Bjorkman, P.J.; Strominger, J.L.; Wiley, D.C.
#journal      J. Mol. Biol. (1985) 186:205
#title        Crystallization and x-Ray diffraction studies on the
              histocompatibility antigens HLA-A2 and HLA-A28 from human
              cell membranes.
COMMENT        Resolution: 2.5 angstroms
COMMENT        Determination: X-ray diffraction
FEATURE        46-47,31-37,21-28,
              3-12,94-103,
              109-118,121-126,
              133-135
              49-53
              57-84
              140-149
              152-161
              163-174
              176-179
              186-195,198-208,
              241-250,229-230
              186-195,198-208,
              241-250,234-235
              222-223,214-219,
              257-262,270-273
              101-164
              203-259
SUMMARY        #length 275 #molecular-weight 31808 #checksum 4235
Query Match    80.0%; Score 64; DB 5; Length 275;
Best Local Similarity 88.9%; Pred. No. 1.19e+00;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0
Db 209 YPAEITLW 217
OY 1 YPAEITLW 9
RESULT 11
ENTRY TITLE 1HHA #type complete
ORGANISM MHC class I histocompatibility antigen HLA-A0201 (with
           nonameric peptide from HTV-1 reverse transcriptase,
           residues 309-317), chain A - human
           #formal_name Homo sapiens #common_name man
           recombinant form expressed in Escherichia coli
           A51226
REFERENCE #authors Madden, D.R.; Garboczi, D.N.; Wiley, D.C.
           #submission Submitted to the Brookhaven Protein Data Bank, June 1993

```

```

#cross-references PDB:1HHJ
REFERENCE
#authors #journal
#title
TN019813
Garbocki, D.N.; Hung, D.T.; Wiley, D.C.
Proc. Natl. Acad. Sci. U.S.A. (1997) 89:3429
HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in escherichia coli and complexed with single antigenic peptides.
TN019814
Tsomides, T.J.; Walker, B.D.; Eisen, H.N.
Proc. Natl. Acad. Sci. U.S.A. (1991) 88:11276
An optimal viral peptide recognized by cd8+ t cells binds very tightly to the restricting class I major histocompatibility complex protein on intact cells but no to the purified class I protein.
A43338
Madden, D.R.; Gorga, J.C.; Strominger, J.L.; Wiley, D.C.
Cell (1992) 70:1035-1048
The three-dimensional structure of HLA-B27 at 2.1 angstrom resolution suggests a general mechanism for tight peptide binding to MHC.
#cross-references MUID:92405152
REFERENCE
#authors #journal
#title
A58495
Saper, M.A.; Bjorkman, P.J.; Wiley, D.C.
J. Mol. Biol. (1991) 219:277
Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 angstroms resolution.
#cross-references MUID:91245570
REFERENCE
#authors #journal
#title
TN019816
Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C.
Nature (1987) 329:506
Structure of the human class I histocompatibility antigen, HLA-A2.
TN019817
Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C.
Nature (1987) 329:512
The foreign antigen binding site and t cell recognition regions of class I histocompatibility antigens.
TN019818
Bjorkman, P.J.; Strominger, J.L.; Wiley, D.C.
J. Mol. Biol. (1985) 186:205
Crystallization and x-ray diffraction studies on the histocompatibility antigens HLA-A2 and hla-A28 from human cell membranes.
COMMENT Resolution: 2.5 angstroms
COMMENT Determination: X-ray diffraction
FEATURE
#resolution 2.5
#determination X-ray diffraction
#length 275 #molecular-weight 31808 #checksum 4235
SUMMARY
Query Match Score 64; DB 5; Length 275; Best Local Similarity 80.9%; Pred. No. 1,19e+00; Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0

```

| RESULT             | 12   |
|--------------------|--|
| ENTRY              | 1H8GA  |
| TITLE              | MHC class I histocompatibility antigen HLA-A*0201 (with nonameric peptide from HIV-1 gp120 envelope protein, residues 195-207), chain A - human  |
| ORGANISM           | #formal_name Homo sapiens #common_name man   |
| REMARKS            | #recombinant form expressed in Escherichia coli A51223   |
| REMARKS            | Madden, D.R.; Garboczi, D.N.; Wiley, D.C. Submitted to the Brookhaven Protein Data Bank, June 1993   |
| REMARKS            | #cross_references PDB:1H8G   |
| REFERENCE          | TN019741   |
| #authors           | Garboczi, D.N.; Hung, D.T.; Wiley, D.C.  |
| #journal           | Proc. Natl. Acad. Sci. U.S.A. (1992) 89:3429   |
| #title             | HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in Escherichia coli and complexed with single antigenic peptides. |
| REFERENCE          | A43338   |
| #authors           | Madden, D.R.; Gorga, J.C.; Strominger, J.L.; Wiley, D.C.   |
| #journal           | Cell (1992) 70:1035-1048   |
| #title             | The three-dimensional structure of HLA-B*27 at 2.1 angstrom resolution suggests a general mechanism for tight peptide binding to MHC.            |
| REFERENCE          | #cross_references M01D:92405152  |
| REFERENCE          | A58495   |
| #authors           | Saper, M.A.; Bjorkman, P.J.; Wiley, D.C.   |
| #journal           | J. Mol. Biol. (1991) 219:277   |
| #title             | Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 angstroms resolution.  |
| REFERENCE          | #cross_references M01D:91245570  |
| REFERENCE          | TN019743   |
| #authors           | Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C.  |
| #journal           | Nature (1987) 329:506  |
| #title             | Structure of the human class I histocompatibility antigen, HLA-A2.   |
| REFERENCE          | TN019744   |
| #authors           | Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C.  |
| #journal           | Nature (1987) 329:512  |
| #title             | The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens.  |
| REFERENCE          | TN019745   |
| #authors           | Bjorkman, P.J.; Strominger, J.L.; Wiley, D.C.  |
| #journal           | J. Mol. Biol. (1985) 186:205   |
| #title             | Crystallization and x-ray diffraction studies on the histocompatibility antigens HLA-A2 and HLA-A28 from human cell membranes.                   |
| COMMENT            | Resolution: 2.6 angstroms  |
| COMMENT            | Determination: x-ray diffraction   |
| FEATURE            |  |
| 46-47,31-37,21-28, |  |
| 3-12,94-103,       |  |
| 109-118,121-126,   |  |
| 133-135            |  |
| 49-53              | #region beta sheet\  |
| 57-84              | #region helix (right hand alpha)\  |
| 140-149            | #region helix (right hand alpha)\  |
| 152-161            | #region helix (right hand alpha)\  |
| 163-174            | #region helix (right hand alpha)\  |
| 176-179            | #region helix (right hand alpha)\  |
| 186-195,198-208,   |  |
| 241-250,223-230,   | #region beta sheet\  |
| 186-195,198-208,   |  |
| 241-250,234-235,   | #region beta sheet\  |
| 222-223,214-219,   |  |
| 257-262,270-273,   | #region beta sheet\  |
| 101-164            | #region beta sheet\  |
| 203-259            | #disulfide_bonds   |





**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
 W O R K I N G  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Merch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 12:50:58 1998; Maspar time 2.38 Seconds

Tabular output not generated. 94.892 Million cell updates/sec

Title: >US-08-452-843-4  
 Description: (1-9) from US08452843.pep  
 Perfect Score: 80  
 Sequence: 1 YPAEITLW 9

Scoring table: PAM 150  
 Gap 15

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot35  
 1:swiss1

Statistics: Mean 25.532; Variance 33.977; scale 0.751

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                        | Pred. No. |
|------------|-------|-------------|--------|-------|------------------------------------|-----------|
| 1          | 64    | 80.0        | 270    | 1     | 1B33_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 2          | 64    | 80.0        | 273    | 1     | 1A69_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 3          | 64    | 80.0        | 275    | 1     | 1A9X_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 4          | 64    | 80.0        | 338    | 1     | 1B20_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 5          | 64    | 80.0        | 354    | 1     | 1B24_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 6          | 64    | 80.0        | 358    | 1     | HLAE_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 7          | 64    | 80.0        | 359    | 1     | HLAE_PONPY CLASS I HISTOCOMPARTIBI | 5.00e-02  |
| 8          | 64    | 80.0        | 359    | 1     | HLAE_PANTR CHLA CLASS I HISTOCOMPA | 5.00e-02  |
| 9          | 64    | 80.0        | 361    | 1     | HLA19_HUMAN HLA CLASS I HISTOCOMPA | 5.00e-02  |
| 10         | 64    | 80.0        | 362    | 1     | HLA19_CANFA DLA CLASS I HISTOCOMPA | 5.00e-02  |
| 11         | 64    | 80.0        | 362    | 1     | HLAF_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 12         | 64    | 80.0        | 362    | 1     | HLAF_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 13         | 64    | 80.0        | 362    | 1     | 1B01_GORGO CLASS I HISTOCOMPARTIBI | 5.00e-02  |
| 14         | 64    | 80.0        | 362    | 1     | 1B02_GORGO CHLA CLASS I HISTOCOMPA | 5.00e-02  |
| 15         | 64    | 80.0        | 362    | 1     | 1B03_GORGO CLASS I HISTOCOMPARTIBI | 5.00e-02  |
| 16         | 64    | 80.0        | 362    | 1     | 1B02_GORGO CLASS I HISTOCOMPARTIBI | 5.00e-02  |
| 17         | 64    | 80.0        | 362    | 1     | 1B02_GORGO CLASS I HISTOCOMPARTIBI | 5.00e-02  |
| 18         | 64    | 80.0        | 362    | 1     | 1B08_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 19         | 64    | 80.0        | 362    | 1     | 1B10_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 20         | 64    | 80.0        | 362    | 1     | 1B11_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 21         | 64    | 80.0        | 362    | 1     | 1B12_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 22         | 64    | 80.0        | 362    | 1     | 1B04_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |
| 23         | 64    | 80.0        | 362    | 1     | 1B07_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02  |

## ALIGNMENTS

|    |    |      |     |   |                                    |          |
|----|----|------|-----|---|------------------------------------|----------|
| 24 | 64 | 80.0 | 362 | 1 | 1B05_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 25 | 64 | 80.0 | 362 | 1 | 1B21_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 26 | 64 | 80.0 | 362 | 1 | 1B18_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 27 | 64 | 80.0 | 362 | 1 | 1B19_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 28 | 64 | 80.0 | 362 | 1 | 1B16_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 29 | 64 | 80.0 | 362 | 1 | 1B15_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 30 | 64 | 80.0 | 362 | 1 | 1B13_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 31 | 64 | 80.0 | 362 | 1 | 1B34_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 32 | 64 | 80.0 | 362 | 1 | 1B32_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 33 | 64 | 80.0 | 362 | 1 | 1B31_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 34 | 64 | 80.0 | 362 | 1 | 1B38_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 35 | 64 | 80.0 | 362 | 1 | 1B39_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 36 | 64 | 80.0 | 362 | 1 | 1B35_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 37 | 64 | 80.0 | 362 | 1 | 1B36_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 38 | 64 | 80.0 | 362 | 1 | 1B23_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 39 | 64 | 80.0 | 362 | 1 | 1B22_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 40 | 64 | 80.0 | 362 | 1 | 1B25_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 41 | 64 | 80.0 | 362 | 1 | 1B26_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 42 | 64 | 80.0 | 363 | 1 | 1B04_GORGO CLASS I HISTOCOMPARTIBI | 5.00e-02 |
| 43 | 64 | 80.0 | 365 | 1 | 1A68_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 44 | 64 | 80.0 | 365 | 1 | 1A74_HUMAN HLA CLASS I HISTOCOMPA  | 5.00e-02 |
| 45 | 64 | 80.0 | 365 | 1 | 1B01_SAGOE CLASS I HISTOCOMPARTIBI | 5.00e-02 |

|             |  |              |           |            |             |
|-------------|--|--------------|-----------|------------|-------------|
| RESULT 1    | ID   | 1B33_HUMAN   | STANDARD: | PRT:       | 270 AA.     |
| AC          | P01890:  |              |           |            |             |
| DT          | 21-JUL-1986 (REL. 01, CREATED)   |              |           |            |             |
| DT          | 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  |              |           |            |             |
| DT          | 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  |              |           |            |             |
| DE          | HLA CLASS I HISTOCOMPARIBILITY ANTIGEN, BW-60(B-40) B*4001 ALPHA CHAIN (FRAGMENT).   |              |           |            |             |
| GN          | HLA-B OR HLAB.   |              |           |            |             |
| OS          | HOMO SAPIENS (HUMAN).  |              |           |            |             |
| OC          | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;                       |              |           |            |             |
| CC          | EUTHERIA; PRIMATES.  |              |           |            |             |
| CC          | [1]  |              |           |            |             |
| CC          | SEQUENCE.  |              |           |            |             |
| CC          | RP MEDLINE; 84000412.  |              |           |            |             |
| CC          | RA LOPEZ DE CASTRO J.A., BRAGADO R., STRONG D.M., STROMINGER J.L.;                   |              |           |            |             |
| CC          | RL BIOCHEMISTRY 22:3961-3969(1983).  |              |           |            |             |
| CC          | -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM. |              |           |            |             |
| CC          | -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).           |              |           |            |             |
| CC          | CC PIR: A02186; HLH040.  |              |           |            |             |
| CC          | DR HSP: P03989; IHS.   |              |           |            |             |
| CC          | DR MIM: 142830; -  |              |           |            |             |
| CC          | DR PROSITE; PS00290; IG_MHC; 1.  |              |           |            |             |
| CC          | CC MHC I; GLYCOPROTEIN.  |              |           |            |             |
| CC          | CC DOMAIN 1  |              |           |            |             |
| CC          | FT DOMAIN 91 181   |              |           |            |             |
| CC          | FT DOMAIN 182 >270   |              |           |            |             |
| CC          | FT CARBOHYD 86 86  |              |           |            |             |
| CC          | FT DISULFID 101 163  |              |           |            |             |
| CC          | FT DISULFID 202 258  |              |           |            |             |
| CC          | FT NON_TER 270 270   |              |           |            |             |
| CC          | CC SEQUENCE 270 AA; 31205 MW; BFE44EFF CRC32;  |              |           |            |             |
| Query Match | Best Local Similarity  | 80.0%;       | Score 64; | DB 1;      | Length 270; |
| Matches     | 8;   | Conservative | 0;        | Mismatches | 1;          |
| Indels      | 0;   | Gaps         | 0;        |            |             |
| DB          | 208 YPAEITLW 216   |              |           |            |             |
| OY          | 1 YPAEITLW 9   |              |           |            |             |
| RESULT 2    | ID   | 1A69_HUMAN   | STANDARD: | PRT:       | 273 AA.     |
| AC          | P10316:  |              |           |            |             |

01-MAR-1989 (REL. 10, CREATED)  
 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, AW-69(A-28) ALPHA CHAIN  
 (FRAGMENT).  
 HLA-A OR HLA.  
 HOMO SAPIENS (HUMAN).  
 EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 [1]  
 SEQUENCE FROM N.A. (A\*6901).  
 MEDLINE; 86055720.  
 RA HOLMES N., PARHAM P.;  
 EMBL; 4:2849-2854(1985).  
 -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 THE IMMUNE SYSTEM.  
 -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 MICROGLOBULIN).  
 -1- POLYMORPHISM: THE ONLY ALLELE OF AW-69 KNOWN IS A\*6901 WHICH IS  
 SHOWN HERE.  
 EMBL; X03158; -; NOT ANNOTATED\_CDS.  
 DR EMBL; X03159; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; B24671; HLH069.  
 DR HSSP; P01892; IHG.  
 DR MIM; 142800; -.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 KM MHC I; TRANSMEMBRANE; GLYCOPROTEIN.  
 FT NON\_TER 1  
 FT DOMAIN 1  
 FT DOMAIN 1  
 FT DOMAIN 90  
 FT DOMAIN 180  
 FT DOMAIN 181  
 FT CARBOHYD 85  
 FT CARBOHYD 85  
 FT DISULFID 100  
 FT DISULFID 163  
 FT NON\_TER 202  
 FT NON\_TER 258  
 FT SEQUENCE 273 AA; 31677 MW; EEBFB366 CRC32;  
 Query Match 80.0%; Score 64; DB 1; Length 273;  
 Best Local Similarity 88.9%; Pred. No. 5.00e-02;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

208 YPAEITLW 216  
 1 YPAEITLW 9

RESULT 3  
 ID 1AXX\_HUMAN STANDARD; PRT; 275 AA.  
 AC P10313;  
 DT 01-MAR-1989 (REL. 10, CREATED)  
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-10 ALPHA CHAIN (FRAGMENT).  
 HLA-A OR HLA.  
 HOMO SAPIENS (HUMAN).  
 EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE; 86033791.  
 RA DAVIDSON W.F., KRESS M., KHOURY G., JAY G.;  
 J. BIOL. CHEM. 260:13414-13423(1985).  
 -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 THE IMMUNE SYSTEM.  
 -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 MICROGLOBULIN).  
 EMBL; M1887; G184158; -.  
 DR PIR; B24512; HLH010.  
 DR HSSP; P01892; IHG.  
 DR MIM; 142800; -.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 KM MHC I; TRANSMEMBRANE; GLYCOPROTEIN.  
 FT NON\_TER 1

FT DOMAIN <1 24  
 FT DOMAIN 25 116  
 FT DOMAIN 117 208  
 FT DOMAIN 209 218  
 FT TRANSMEM 219 242  
 FT DOMAIN 243 275  
 FT CARBOHYD 20 20  
 FT DISULFID 35 98  
 FT DISULFID 137 193  
 FT SEQUENCE 273 AA; 30548 MW; 8B232F3C CRC32;  
 Query Match 80.0%; Score 64; DB 1; Length 275;  
 Best Local Similarity 88.9%; Pred. No. 5.00e-02;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

143 YPAEITLW 151  
 1 YPAEITLW 9

RESULT 4  
 ID 1B2A\_HUMAN STANDARD; PRT; 354 AA.  
 AC P30467;  
 DT 01-APR-1993 (REL. 25, CREATED)  
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B\*2707 ALPHA CHAIN  
 (B27-HS).  
 HLA-B OR HLAB.  
 HOMO SAPIENS (HUMAN).  
 EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE; 91268545.  
 RA CHOO Y.S., FAN L.A., HANSEN J.A.;  
 J. IMMUNOL. 147:174-180(1991).  
 -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 THE IMMUNE SYSTEM.  
 -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 MICROGLOBULIN).  
 EMBL; M62852; G187761; -.  
 DR HSSP; P03989; IHS.  
 DR MIM; 142830; -.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 KM MHC I; TRANSMEMBRANE; GLYCOPROTEIN.  
 FT DOMAIN 1  
 FT DOMAIN 90  
 FT DOMAIN 91  
 FT DOMAIN 182  
 FT DOMAIN 183  
 FT DOMAIN 274  
 FT DOMAIN 275  
 FT TRANSMEM 285 308  
 FT TRANSMEM 309 338  
 FT CARBOHYD 86 86  
 FT DISULFID 101 164  
 FT DISULFID 203 259  
 FT SEQUENCE 338 AA; 37804 MW; 33FB8134 CRC32;  
 Query Match 80.0%; Score 64; DB 1; Length 338;  
 Best Local Similarity 88.9%; Pred. No. 5.00e-02;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

209 YPAEITLW 217  
 1 YPAEITLW 9

RESULT 5  
 ID 1B2A\_HUMAN STANDARD; PRT; 354 AA.  
 AC P30470;  
 DT 01-APR-1993 (REL. 25, CREATED)  
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-35 B\*3504 ALPHA CHAIN

```

DE PRECURSOR (FRAGMENT).
GN HLA-B OR HLAB.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92269956.
RA WATKINS D.I., MCADAM S.N., LIU X., STANG C.R., MILFORD E.L.,
RA LEVINE C.G., GARBER T.L., DOSON A.L., LORD C.I., GHIM S.H.,
RA TROUP G.M., HUGHES A.L., LEVIN N.L.;
RL NATURE 357:329-333(1992).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
DR EMBL: M86403; -; NOT_ANNOTATED_CDS.
DR HSSP: P03989; 1HSA.
DR MIM: 142830; -.
DR PROSITE: PS00290; IG_MHC; 1.
DR MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT FT NON_TER 1 1
FT SIGNAL <1 16
FT CHAIN 17 354
FT FT 17 106 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 107 198 ALPHA CHAIN B-35 B*3504.
FT DOMAIN 109 290 EXTRACELLULAR ALPHA-1.
FT DOMAIN 291 300 EXTRACELLULAR ALPHA-2.
FT TRANSMEM 301 324 EXTRACELLULAR ALPHA-3.
FT DOMAIN 325 354 CONNECTING PEPTIDE.
FT CARBOHYD 102 102 CYTOPLASMIC TAIL.
FT DISULFID 117 180 BY SIMILARITY.
FT DISULFID 219 275 BY SIMILARITY.
FT DISULFID 219 275 BY SIMILARITY.
SQ SEQUENCE 354 AA: 39617 MW: 6564795A CRC32:

Query Match
Best Local Similarity 80.0%; Score 64; DB 1; Length 354;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 225 YPAEITLW 233
QY 1 YPAEITLW 9

RESULT 6
ID HLAE_HUMAN STANDARD; PRT; 358 AA.
AC P13747;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, E*0101/E*0102 ALPHA CHAIN
DE PRECURSOR.
GN HLA-B OR HLAB-6.2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A. (E*0101).
RX MEDLINE: 88229102.
RA MIZUNO S., TRAPANI J.A., KOLLER B.H., DUPONT B., YANG S.Y.;
RL J. IMMUNOL. 140:4024-4030(1988).
RN [2]
RP SEQUENCE FROM N.A. (E*0102).
RX MEDLINE: 88285691.
RA KOLLER B.H., GERAGHTY D.E., SHIMIZU Y., DEMARS R., ORR H.T.;
RL J. IMMUNOL. 141:897-904(1988).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -1- THE SEQUENCE SHOWN IS THAT OF E*0101.
DR EMBL: M20022; G306852; -

```

```

DR EMBL: M21533; G386948; -
DR PIR: A32272; A32272.
DR PIR: A28834; A28834.
DR HSSP: P03989; 1HSA.
DR MIM: 143010; -.
DR PROSITE: PS00290; IG_MHC; 1.
DR MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT FT SIGNAL 1 21
FT CHAIN 22 358
FT FT 22 111 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 112 203 ALPHA CHAIN E*0101/E*0102.
FT DOMAIN 204 295 EXTRACELLULAR ALPHA-1.
FT DOMAIN 296 305 EXTRACELLULAR ALPHA-2.
FT TRANSMEM 306 329 EXTRACELLULAR ALPHA-3.
FT DOMAIN 330 358 CONNECTING PEPTIDE.
FT DISULFID 122 185 CYTOPLASMIC TAIL.
FT DISULFID 224 280 BY SIMILARITY.
FT CARBOHYD 107 107 BY SIMILARITY.
FT CONFLICT 10 10 S -> L (IN E*0102).
FT CONFLICT 104 104 G -> R (IN E*0102).
SQ SEQUENCE 358 AA: 40130 MW: 3D79F233 CRC32:

Query Match
Best Local Similarity 80.0%; Score 64; DB 1; Length 358;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 230 YPAEITLW 238
QY 1 YPAEITLW 9

RESULT 7
ID HLAE_PONPY STANDARD; PRT; 359 AA.
AC P16212;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE CLASS I HISTOCOMPATIBILITY ANTIGEN, E-1 ALPHA CHAIN PRECURSOR
DE (FRAGMENT).
GN PONGO PYGMAEUS (ORANGUTAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90201944.
RA LAWLER D.A., WARREN E., PARHAM P.;
RL IMMUNOL. REV. 113:147-185(1990).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
DR EMBL: M30681; G342846; -.
DR HSSP: P03989; 1HSA.
DR PROSITE: PS00290; IG_MHC; 1.
DR MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT FT NON_TER 1 1
FT SIGNAL <1 18
FT CHAIN 19 359
FT FT 19 108 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 109 200 E-1 ALPHA CHAIN.
FT DOMAIN 201 292 EXTRACELLULAR ALPHA-1.
FT DOMAIN 293 302 EXTRACELLULAR ALPHA-2.
FT TRANSMEM 303 326 EXTRACELLULAR ALPHA-3.
FT DOMAIN 327 359 CONNECTING PEPTIDE.
FT DISULFID 119 182 CYTOPLASMIC TAIL.
FT DISULFID 221 277 BY SIMILARITY.
FT CARBOHYD 104 104 BY SIMILARITY.
SQ SEQUENCE 359 AA: 40409 MW: 55E15638 CRC32:

Query Match
Best Local Similarity 80.0%; Score 64; DB 1; Length 359;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 227 YPAEITLW 235  
 1 YPAEITLW 9

RESULT 8  
 ID 1B01.PANTR STANDARD: PRT: 359 AA.  
 AC P13750;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
 DE CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-1 ALPHA CHAIN PRECURSOR  
 DI (FRAGMENT).  
 DS PAN TROGLODYTES (CHIMPANZEE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RX MEDLINE: 89030641.  
 RA KLEIN J. J.; JONKER M., KLEIN D., IVANYI P., VAN SEVENTER G.,  
 RA EMBO J. 7:2765-2774(1988).  
 RN [2]  
 RP REVISIONS.  
 RA SUBMITTED (FEB-1989) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC EMBL: X13115; G755776;  
 DR PIR: S03537; S03837.  
 DR HSSP: P03989; ILSA.  
 DR PROSITE: P500290; IG\_MHC; 1.  
 DR MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.  
 FT NON TER 1  
 FT SIGNAL <1 20  
 FT CHAIN 21 359  
 FT 1  
 FT DOMAIN 21 110 CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT DOMAIN 111 110 B-1 ALPHA CHAIN.  
 FT DOMAIN 203 294 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 285 305 EXTRACELLULAR ALPHA-2.  
 FT TRANSMEM 306 329 CONNECTING PEPTIDE.  
 FT DOMAIN 330 359 CYTOPLASMIC TAIL.  
 FT DISULFID 121 184 BY SIMILARITY.  
 FT DISULFID 223 279 BY SIMILARITY.  
 FT CARBOHYD 106 106 BY SIMILARITY.  
 SO SEQUENCE 359 AA; 40173 MW; 5395FEC9 CRC32;

Query Match 80.0%; Score 64; DB 1; Length 359;  
 Best Local Similarity 88.9%; Pred. No. 5.00e-02;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 229 YPAEITLW 237  
 1 YPAEITLW 9

RESULT 9  
 ID 1B14.HUMAN STANDARD: PRT: 361 AA.  
 AC P03989;  
 DT 23-OCT-1986 (REL. 02, CREATED)  
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
 DE HLA-B OR HLAB.  
 GN HLA-B OR HLAB.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE: 86138405.  
 RA WEISS E.H., KUON W., DOERNER C., LANG M., RIETHMUELLER G.;  
 RL IMMUNOBIOLOGY 170:367-380(1985).  
 RN [2]  
 RP SEQUENCE OF 25-361 FROM N.A.  
 RX MEDLINE: 86149317.  
 RA SZOETS H., RIETHMUELLER G., WEISS E., MEO T.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:1428-1432(1986).  
 RN [3]  
 RP SEQUENCE OF 25-295.  
 RX MEDLINE: 85226361.  
 RA EZQUERRA A., BRAGADA R., VEGA M.A., STROMINGER J.L., WOODY J.,  
 RA LOPEZ DE CASTRO J.A.;  
 RL BIOCHEMISTRY 24:1733-1741(1985).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 25-300.  
 RX MEDLINE: 92405152.  
 RA MADDEN D.R., GORGA J.C., STROMINGER J.L., WILEY D.C.;  
 RL CELL 70:1035-1048(1992).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY.  
 RX MEDLINE: 92018187.  
 RA MADDEN D.R., GORGA J.C., STROMINGER J.L., WILEY D.C.;  
 RL NATURE 353:321-325(1991).  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN).  
 CC -1- DISEASE: THIS PROTEIN CORRELATES WITH THE DEVELOPMENT OF  
 CC ANKYLOSING SPONDYLITIS.  
 CC EMBL: X03945; G32177; ALT\_TERM.  
 DR PIR: A25128; H1HUB2.  
 DR PIR: S07441; S07441.  
 DR PDB: ILSA; 15-OCT-92.  
 DR MIM: 142830;  
 DR PROSITE: P500290; IG\_MHC; 1.  
 KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL; 3D-STRUCTURE.  
 FT SIGNAL 1 24  
 FT CHAIN 25 361  
 FT 1  
 FT DOMAIN 25 114 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 FT DOMAIN 115 114 ALPHA CHAIN B-27.  
 FT DOMAIN 207 298 EXTRACELLULAR ALPHA-1.  
 FT DOMAIN 299 308 EXTRACELLULAR ALPHA-2.  
 FT TRANSMEM 309 332 CONNECTING PEPTIDE.  
 FT DOMAIN 333 361 CYTOPLASMIC TAIL.  
 FT CARBOHYD 110 110  
 FT DISULFID 125 188  
 FT DISULFID 227 283  
 FT CONFLICT 206 206  
 FT CONFLICT 266 266  
 FT STRAND 27 38  
 FT TURN 39 41  
 FT STRAND 42 52  
 FT TURN 53 54  
 FT STRAND 55 61  
 FT TURN 62 63  
 FT STRAND 70 71  
 FT TURN 74 76  
 FT HELIX 77 78  
 FT TURN 81 108  
 FT HELIX 109 110  
 FT TURN 113 114  
 FT STRAND 118 127  
 FT TURN 129 130  
 FT STRAND 133 142  
 FT TURN 143 144  
 FT STRAND 145 150  
 FT TURN 152 153  
 FT STRAND 157 159  
 FT HELIX 162 173  
 FT TURN 174 175

A -> V (IN REF. 2).  
 Q -> E (IN REF. 3).

```

FT HELIX 176 185
FT TURN 186 186
FT HELIX 187 198
FT TURN 189 199
FT HELIX 200 203
FT TURN 204 204
FT STRAND 207 207
FT STRAND 210 217
FT STRAND 222 233
FT STRAND 238 243
FT TURN 244 245
FT STRAND 246 247
FT HELIX 249 251
FT STRAND 253 254
FT STRAND 258 259
FT STRAND 265 274
FT TURN 275 276
FT HELIX 278 280
FT STRAND 281 286
FT TURN 288 289
FT STRAND 294 296
SO SEQUENCE 361 AA; 40462 MM; 802130D5 CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 1; Length 361;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
|||||||
QY 1 YPAEITLW 9

RESULT 10
ID HA19_CANFA STANDARD; PRT; 362 AA.
AC P18466;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
DE DLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A9/A9 ALPHA CHAIN PRECURSOR.
OS CANIS FAMILIARIS (DOG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; CARNIVORA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90316611.
RA SARMENTO U.M., STORR R.;
RL IMMUNOGENETICS 31:400-404(1990).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC EMBL: M32283; G164005; -
DR PIR: A45845; A45845.
DR HSSP: P03989; ILSA.
DR PROSITE: PS00290; IG_MHC.1.
KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 24
FT CHAIN 25 362
FT DOMAIN 25 114
FT DOMAIN 115 207
FT DOMAIN 208 299
FT DOMAIN 300 306
FT TRANSMEM 307 329
FT DOMAIN 330 362
FT DISULFID 125 189
FT DISULFID 228 284
FT CARBOHYD 110 110
FT SEQUENCE 362 AA; 40462 MM; D5250E8D CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 1; Length 362;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 234 YPAEITLW 242
|||||||
QY 1 YPAEITLW 9

RESULT 11
ID HLAH_HUMAN STANDARD; PRT; 362 AA.
AC P01933;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, H ALPHA CHAIN PRECURSOR
DE (HLA-A*01:01).
GN HLA-A OR HLA-A.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 82151002.
RA MALISSEN M., MALISSEN B., JORDAN B.R.;
RL PROC. NATL. ACAD. SCI. U.S.A. 79:893-897(1982).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM. COULD BE THE PRODUCT OF A PSEUDOGENE.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
DR EMBL: J00191; G386873; ALT_INIT.
DR PIR: A02189; HLH012.
DR HSSP: P03989; ILSA.
DR MIM: 142925; -
DR PROSITE: PS00290; IG_MHC.1.
KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 24
FT CHAIN 25 362
FT DOMAIN 25 114
FT DOMAIN 115 207
FT DOMAIN 207 298
FT DOMAIN 299 308
FT TRANSMEM 309 332
FT DOMAIN 333 362
FT CARBOHYD 110 110
FT DISULFID 227 283
FT SEQUENCE 362 AA; 40850 MM; 5E610F63 CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 1; Length 362;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
|||||||
QY 1 YPAEITLW 9

RESULT 12
ID HLAH_HUMAN STANDARD; PRT; 362 AA.
AC P30511;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, F ALPHA CHAIN PRECURSOR (HLA F
DE ANTIGEN) (LEUCOCYTE ANTIGEN F) (CD12).
GN HLA-F OR HLA-F OR HLA-5.4.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90111605.
RA GERAGHTY D.E., WEI X., ORR H.T., KOLLER B.H.;
RL J. EXP. MED. 171:1-18(1990).
RN [2]

```

RP SEQUENCE FROM N.A.  
RA MEDLINE: 9197889.  
RA LURY D., EPSTEIN H., HOLMES N.:  
RL INT. IMMUNOL. 2:531-537(1990).  
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
CC THE IMMUNE SYSTEM.  
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN).  
CC EMBL: X17093; G312407.  
DR PIR: A60384; A60384.  
DR PIR: JLO147; JLO147.  
DR HSSP: P03989; 1HSA.  
DR MIM: 143110.  
DR PROSITE: PS00290; IG\_MHC; 1.  
KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.  
FT SIGNAL 1 21  
FT CHAIN 22 362  
FT FT  
FT DOMAIN 22 111 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
FT DOMAIN 112 203 ALPHA CHAIN F.  
FT DOMAIN 204 295 EXTRACELLULAR ALPHA-1.  
FT DOMAIN 296 305 EXTRACELLULAR ALPHA-2.  
FT TRANSMEM 306 329 CONNECTING PEPTIDE.  
FT DOMAIN 330 362 CYTOPLASMIC TAIL.  
FT DISULFID 122 185 BY SIMILARITY.  
FT DISULFID 224 280 BY SIMILARITY.  
FT CARBOHYD 107 107 BY SIMILARITY.  
SQ SEQUENCE 362 AA; 40568 MW; E9B29521 CRC32;  
Query Match 80.0%; Score 64; DB 1; Length 362;  
Best Local Similarity 88.9%; Pred. No. 5.00e-02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 230 YPAEITLW 238  
| | | | | | | | | |  
QY 1 YPAEITLW 9.  
RESULT 13  
ID 1801 GORGO STANDARD; PRT; 362 AA.  
AC P30379;  
DT 01-APR-1993 (REL. 25; CREATED)  
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
DE 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
DE CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-B0101 ALPHA CHAIN PRECURSOR.  
OS GORILLA GORILLA GORILLA (LOWLAND GORILLA).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUKARYOTA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE: 92078860.  
RA LAMTOR D.A., WARREN E., TAYLOR P., PARHAM P.;  
RL J. EXP. MED. 174:1491-1509(1991).  
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
CC THE IMMUNE SYSTEM.  
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN).  
CC EMBL: X60355; G32866;  
DR PIR: JH0539; JH0539.  
DR HSSP: P03989; 1HSA.  
DR PROSITE: PS00290; IG\_MHC; 1.  
KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.  
FT SIGNAL 1 24  
FT CHAIN 25 362  
FT FT  
FT DOMAIN 25 114 GOGO-B0101 ALPHA CHAIN.  
FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.  
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.  
FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.  
FT TRANSMEM 309 332 CONNECTING PEPTIDE.  
FT DOMAIN 333 362 CYTOPLASMIC TAIL.  
FT DISULFID 125 188 BY SIMILARITY.  
FT DISULFID 227 283 BY SIMILARITY.  
FT CARBOHYD 110 110 BY SIMILARITY.  
SQ SEQUENCE 362 AA; 40488 MW; 4BF65A6C CRC32;  
Query Match 80.0%; Score 64; DB 1; Length 362;  
Best Local Similarity 88.9%; Pred. No. 5.00e-02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 233 YPAEITLW 241  
| | | | | | | | | |  
QY 1 YPAEITLW 9

FT CARBOHYD 110 110 BY SIMILARITY.  
SQ SEQUENCE 362 AA; 40170 MW; 2E33E2B8 CRC32;  
Query Match 80.0%; Score 64; DB 1; Length 362;  
Best Local Similarity 88.9%; Pred. No. 5.00e-02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 233 YPAEITLW 241  
| | | | | | | | | |  
QY 1 YPAEITLW 9  
RESULT 14  
ID 1802 PANTR STANDARD; PRT; 362 AA.  
AC P13731;  
DT 01-JAN-1990 (REL. 13, CREATED)  
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
DE 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
DE CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-2 ALPHA CHAIN PRECURSOR.  
OS PAN TROGLODITES (CHIMPANZEE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUKARYOTA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE: 89030641.  
RA MAYER W.E., JONKER M., KLEIN D., IYANYI P., VAN SEVENTER G.,  
RA KLEIN J.,  
RL EMBO J. 7:2765-2774(1988).  
RN [2]  
RP REVISIONS.  
RA MAYER W.;  
RL SUBMITTED (FEB-1989) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
CC THE IMMUNE SYSTEM.  
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN).  
CC EMBL: X13116; G38209;  
DR PIR: S03538; S03538.  
DR HSSP: P03989; 1HSA.  
DR PROSITE: PS00290; IG\_MHC; 1.  
KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.  
FT SIGNAL 1 24  
FT CHAIN 25 362  
FT FT  
FT DOMAIN 25 114 CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
FT DOMAIN 115 206 B-2 ALPHA CHAIN  
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-1.  
FT DOMAIN 299 308 EXTRACELLULAR ALPHA-2.  
FT TRANSMEM 309 332 CONNECTING PEPTIDE.  
FT DOMAIN 333 362 CYTOPLASMIC TAIL.  
FT DISULFID 125 188 BY SIMILARITY.  
FT DISULFID 227 283 BY SIMILARITY.  
FT CARBOHYD 110 110 BY SIMILARITY.  
SQ SEQUENCE 362 AA; 40488 MW; 4BF65A6C CRC32;  
Query Match 80.0%; Score 64; DB 1; Length 362;  
Best Local Similarity 88.9%; Pred. No. 5.00e-02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 233 YPAEITLW 241  
| | | | | | | | | |  
QY 1 YPAEITLW 9  
RESULT 15  
ID 1803 GORGO STANDARD; PRT; 362 AA.  
AC P30381;  
DT 01-APR-1993 (REL. 25, CREATED)  
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
DE 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
DE CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-B0103 ALPHA CHAIN PRECURSOR.  
OS GORILLA GORILLA GORILLA (LOWLAND GORILLA).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA: PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92078860.  
 RA LAWLOR D.A., WARREN E., TAYLOR P., PARRAM P.;  
 RL J. EXP. MED. 174:1491-1509(1991).  
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO  
 CC THE IMMUNE SYSTEM.  
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
 CC MICROGLOBULIN)  
 DR EMBL: X60254; G22870; -.  
 DR PIR: JH0541; JH0541.  
 DR HSSP: P03989; IHSB.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 KW MHC I; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 362  
 FT  
 FT DOMAIN 25 114  
 FT DOMAIN 115 206  
 FT DOMAIN 207 298  
 FT DOMAIN 299 308  
 FT DOMAIN 309 332  
 FT DOMAIN 333 362  
 FT DISULFID 125 188  
 FT DISULFID 227 283  
 FT CARBOHYD 110 110  
 SO SEQUENCE 362 AA; 40248 MW; FEA6A941 CRC32;

Query Match 80.0%; Score 64; DB 1; Length 362;  
 Best Local Similarity 88.9%; Pred. No. 5,00e-02;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 233 YPAEITLW 241  
 |||||  
 QY 1 YPAEITLW 9

Search completed: Fri Sep 11 12:51:03 1998  
 Job time: 5 secs.

**THIS PAGE BLANK (USPTO)**



(TM)

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Fri Sep 11 12:51:21 1998;  MasPar time 3.62 Seconds
```

Tabular output not generated.

Title: >US-08-452-843-4  
Description: (1-9) from US08452843.pdf

Sequence: 1 YPAETITLYW 9

Scoring table: PAM 150

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

1:sp\_fungi2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mhc 6:sp\_organelle 7:sp\_phage 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp Vertebrate  
13:sp\_undefined

Statistics: Mean 25.011; Variance 40.992; scale 0.610

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description            | Pred. No. |
|------------|-------|-------------|--------|----|--------|------------------------|-----------|
| 1          | 68    | 85.0        | 1226   | 9  | O56836 | HYPOTHETICAL PROTEIN M | 1.36e-01  |
| 2          | 65    | 81.3        | P79607 | 5  | P79607 | MHC CLASS I PLA-A1 ALP | 4.55e-01  |
| 3          | 65    | 81.3        | 355    | 5  | O30896 | MHC CLASS I PIP1-G*03  | 4.55e-01  |
| 4          | 64    | 80.0        | 225    | 5  | P79542 | MHC CLASS I ANTIGEN (F | 6.77e-01  |
| 5          | 64    | 80.0        | 246    | 5  | O29945 | MHC CLASS I HLA-A CELL | 6.77e-01  |
| 6          | 64    | 80.0        | 298    | 5  | O95350 | HLA-B*1521 (FRAGMENT)  | 6.77e-01  |
| 7          | 64    | 80.0        | 298    | 5  | O19652 | HLA-C*1203 (FRAGMENT)  | 6.77e-01  |
| 8          | 64    | 80.0        | 298    | 5  | O95349 | HLA-B*1528 (FRAGMENT)  | 6.77e-01  |
| 9          | 64    | 80.0        | 298    | 5  | O19653 | HLA-C*0304 (FRAGMENT)  | 6.77e-01  |
| 10         | 64    | 80.0        | 298    | 5  | O28866 | MHC CLASS I HLA-C ALLE | 6.77e-01  |
| 11         | 64    | 80.0        | 298    | 5  | O19655 | HLA-C*0801 (FRAGMENT)  | 6.77e-01  |
| 12         | 64    | 80.0        | 302    | 5  | O95392 | LEUCOCYTE ANTIGEN B (H | 6.77e-01  |
| 13         | 64    | 80.0        | 322    | 5  | O28654 | HLA-B*71 (FRAGMENT)    | 6.77e-01  |
| 14         | 64    | 80.0        | 338    | 5  | O30207 | HLA-B*62 ANTIGEN (FRAG | 6.77e-01  |
| 15         | 64    | 80.0        | 356    | 5  | O30443 | MHC CLASS I CAVA-G*03  | 6.77e-01  |
| 16         | 64    | 80.0        | 356    | 5  | O30441 | MHC CLASS I CAVA-G*01  | 6.77e-01  |
| 17         | 64    | 80.0        | 356    | 5  | O30440 | MHC CLASS I CAVA-G*02  | 6.77e-01  |
| 18         | 64    | 80.0        | 357    | 5  | O30226 | MHC CLASS I AOTR-G*01  | 6.77e-01  |
| 19         | 64    | 80.0        | 357    | 5  | O95427 | MHC CLASS I (FRAGMENT) | 6.77e-01  |
| 20         | 64    | 80.0        | 357    | 5  | O30899 | MHC CLASS I PIP1-G*05  | 6.77e-01  |

|    |      |      |     |        |                        |                         |            |          |
|----|------|------|-----|--------|------------------------|-------------------------|------------|----------|
| 21 | 64   | 80.0 | 357 | 5      | Q30222                 | MHC CLASS I             | ATBE-G-02  | 6.77e-01 |
| 22 | 64   | 80.0 | 350 | 5      | Q30481                 | MHC CLASS I             |            | 6.77e-01 |
| 23 | 64   | 80.0 | 362 | 5      | Q29749                 | MHC CLASS I             | HLA-B HEAV | 6.77e-01 |
| 24 | 80.0 | 362  | 5   | P79456 | HLA-B ALPHA CHAIN (B*5 |                         |            | 6.77e-01 |
| 25 | 64   | 80.0 | 362 | 5      | Q11602                 | MHC CLASS I ANTIGEN     |            | 6.77e-01 |
| 26 | 64   | 80.0 | 362 | 5      | Q29638                 | MHC CLASS I ANTIGEN     |            | 6.77e-01 |
| 27 | 64   | 80.0 | 362 | 5      | Q296981                | MHC HLA-B CELL SURFACE  |            | 6.77e-01 |
| 28 | 64   | 80.0 | 362 | 5      | Q29664                 | MHC CLASS I HLA-B (HLA  |            | 6.77e-01 |
| 29 | 64   | 80.0 | 362 | 5      | Q29538                 | MHC CLASS I LYMPHOCTE   |            | 6.77e-01 |
| 30 | 64   | 80.0 | 362 | 5      | Q29911                 | TRANSMEMBRANE GLYCOPRO  |            | 6.77e-01 |
| 31 | 64   | 80.0 | 362 | 5      | Q29845                 | HLA-B*1802.             |            | 6.77e-01 |
| 32 | 64   | 80.0 | 362 | 5      | Q29844                 | MAJOR HISTOCOMPARTIDILI |            | 6.77e-01 |
| 33 | 64   | 80.0 | 362 | 5      | Q29836                 | HLA-B-6701.             |            | 6.77e-01 |
| 34 | 64   | 80.0 | 363 | 5      | Q29840                 | MHC HLA-A*0301 BLANK G  |            | 6.77e-01 |
| 35 | 64   | 80.0 | 363 | 5      | Q30870                 | MHC CLASS I A.          |            | 6.77e-01 |
| 36 | 64   | 80.0 | 365 | 5      | Q95362                 | LEUCOCYTE ANTIGEN, HLA  |            | 6.77e-01 |
| 37 | 64   | 80.0 | 365 | 5      | Q29907                 | HLA-A*2404 (HLA-A2AK)   |            | 6.77e-01 |
| 38 | 64   | 80.0 | 365 | 5      | Q02939                 | HLA-AW33.1 PRECURSOR.   |            | 6.77e-01 |
| 39 | 64   | 80.0 | 365 | 5      | Q19756                 | MHC CLASS I HLA-A.      |            | 6.77e-01 |
| 40 | 64   | 80.0 | 365 | 5      | Q95352                 | HLA-A*0218.             |            | 6.77e-01 |
| 41 | 64   | 80.0 | 365 | 5      | Q30718                 | MHC CLASS I ANTIGEN MA  |            | 6.77e-01 |
| 42 | 64   | 80.0 | 366 | 5      | Q29652                 | HLA-CW7.                |            | 6.77e-01 |
| 43 | 64   | 80.0 | 366 | 5      | P79461                 | HLA-C ALPHA CHAIN FOR   |            | 6.77e-01 |
| 44 | 64   | 80.0 | 366 | 5      | Q29648                 | MHC CLASS I HISTOCOMPA  |            | 6.77e-01 |
| 45 | 64   | 80.0 | 366 | 5      | Q95603                 | HLA-CW*0702.            |            | 6.77e-01 |

## ALIGNMENTS

|                       |  |                     |           |                    |
|-----------------------|--|---------------------|-----------|--------------------|
| RESULT                | 1  | PRELIMINARY:        | PRT:      | 1226 AA.           |
| ID                    | 058836   |                     |           |                    |
| AC                    | 058836:  |                     |           |                    |
| DT                    | 01-JAN-1998 (TREMBLREL. 05, CREATED)                                   |                     |           |                    |
| DT                    | 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)                      |                     |           |                    |
| DE                    | 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)                    |                     |           |                    |
| DE                    | HYPOTHETICAL PROTEIN MJ1441.   |                     |           |                    |
| GN                    | MJ1441.  |                     |           |                    |
| OS                    | METHANOCOCCUS JANNASCHII.  |                     |           |                    |
| OC                    | ARCHAEBACTERIA: EURYARCHAEOTA: METHANOCOCCALES: METHANOCOCCACEAE.      |                     |           |                    |
| RN                    | [1]  |                     |           |                    |
| RP                    | SEQUENCE FROM N.A.   |                     |           |                    |
| RX                    | MEDLINE: 96337999.   |                     |           |                    |
| RA                    | BOLT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,            |                     |           |                    |
| RA                    | STITTON G.G., BLANK J.A., FITZGERALD L.M., CLAYTON R.A., GOGAINE J.D., |                     |           |                    |
| RA                    | KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,    |                     |           |                    |
| RA                    | OVERBERG R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,   |                     |           |                    |
| RA                    | SCOTT J.L., GEORGHAGEN N.S.M., WEIDMAN J.F., FUHRMAN J.L., NOUYEN D.,  |                     |           |                    |
| RA                    | UTTERBACK T.R., KELEY J.M., PETERSON J.D., SHOWN P.W., HANNA M.C.,     |                     |           |                    |
| RA                    | COTTON M.D., ROBERTS K.M., HORST M.A., KAINE B.P., BORODOVSKY M.,      |                     |           |                    |
| RA                    | KLENN H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.:         |                     |           |                    |
| RL                    | SCIENCE 273:1058-1073(1996).   |                     |           |                    |
| CC                    | -1- SIMILARITY: STRONG TO P.DENTRIFICANS COBN AND M.JANNASCHII         |                     |           |                    |
| CC                    | MJ0907.  |                     |           |                    |
| DR                    | EMBL: U67585; GI500323; -  |                     |           |                    |
| KW                    | HYPOTHETICAL PROTEIN.  |                     |           |                    |
| SQ                    | SEQUENCE 1226 AA; 141327 MW; 4223043D CRC32;                           |                     |           |                    |
| Query Match           |  | 85.0%:              | Score 68; | DB 9; Length 1226; |
| Best Local Similarity | 66.7%:   | Pred. NO. 1.36e-01; |           |                    |
| Matches               | 6; Conservative  | 2; Mismatches       | 1; Indels | 0; Gaps            |
| DB                    | 845 XPENALTYW 853  |                     |           |                    |
| OY                    | 1 YPAEITLYW 9  |                     |           |                    |
| RESULT                | 2  | PRELIMINARY:        | PRT:      | 335 AA.            |
| ID                    | P79607   |                     |           |                    |
| AC                    | P79607:  |                     |           |                    |
| DT                    | 01-MAY-1997 (TREMBLREL. 03, CREATED)                                   |                     |           |                    |
| DT                    | 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)                      |                     |           |                    |
| DE                    | 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)                    |                     |           |                    |
| DE                    | MHC CLASS I PLA-A1 ALPHA-CHAIN (FERAMONT)                              |                     |           |                    |

|        |   |                 |                     |                      |
|--------|---|-----------------|---------------------|----------------------|
| SQL    | SEQUENCE  | 225 AA;         | 25906 MW;           | 136BC712 CRC32;      |
|        | Query Match   | 80.0%;          | Score 64;           | DB 5; Length 225;    |
|        | Best Local Similarity   | 88.9%;          | Pred. No. 6,77e-01; |                      |
|        | Matches   | 8; Conservative | 0; Mismatches       | 1; Indels 0; Gaps 0; |
| Db     | 201 YPAEITLTW 209   |                 |                     |                      |
| QY     | 1 YPAEITLTW 9   |                 |                     |                      |
| RESULT | 5   | PRELIMINARY;    | PRT;                | 246 AA.              |
| ID     | Q29945  |                 |                     |                      |
| AC     | Q29945  |                 |                     |                      |
| DT     | 01-NOV-1996 (TREMBLREL. 01, CREATED)                          |                 |                     |                      |
| DT     | 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)             |                 |                     |                      |
| DT     | 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)           |                 |                     |                      |
| DE     | MHC CLASS I HLA-A CELL SURFACE ANTIGEN (FRAGMENT).            |                 |                     |                      |
| OS     | HOMO SAPIENS (HUMAN)  |                 |                     |                      |
| OC     | EUCAROTA; METAEOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; |                 |                     |                      |
| CC     | EUHERIA; PRIMATES.  |                 |                     |                      |
| RN     | [1]   |                 |                     |                      |
| RP     | SEQUENCE FROM N.A.  |                 |                     |                      |
| RX     | MEDLINE: 84287690.  |                 |                     |                      |
| RA     | ARNOT D., LITTLE J W., AUFFRAY C., KAPPE D., STROMINGER J.L.; |                 |                     |                      |
| RL     | IMMUNOGENTITIS 20:237-252(1984).                              |                 |                     |                      |
| EMBL   | M27537; G386887; -.   |                 |                     |                      |
| DR     | PROSITE: PS00290; IG_MHC; 1.                                  |                 |                     |                      |
| KW     | MHC.  |                 |                     |                      |
| FT     | NON_TER   | 1               |                     |                      |
| SO     | SEQUENCE  | 246 AA;         | 27577 MW;           | A635CFC6 CRC32;      |
|        | Query Match   | 80.0%;          | Score 64;           | DB 5; Length 246;    |
|        | Best Local Similarity   | 88.9%;          | Pred. No. 6,77e-01; |                      |
|        | Matches   | 8; Conservative | 0; Mismatches       | 1; Indels 0; Gaps 0; |
| Db     | 114 YPAEITLTW 122   |                 |                     |                      |
| QY     | 1 YPAEITLTW 9   |                 |                     |                      |
| RESULT | 6   | PRELIMINARY;    | PRT;                | 298 AA.              |
| ID     | Q95350  |                 |                     |                      |
| AC     | Q95350  |                 |                     |                      |
| DT     | 01-FEB-1997 (TREMBLREL. 02, CREATED)                          |                 |                     |                      |
| DT     | 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)             |                 |                     |                      |
| DT     | 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)           |                 |                     |                      |
| DE     | HLA-B*1521 (FRAGMENT).  |                 |                     |                      |
| OS     | HOMO SAPIENS (HUMAN)  |                 |                     |                      |
| OC     | EUCAROTA; METAEOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; |                 |                     |                      |
| CC     | EUHERIA; PRIMATES.  |                 |                     |                      |
| RN     | [1]   |                 |                     |                      |
| RP     | SEQUENCE FROM N.A.  |                 |                     |                      |
| RC     | TISSUE-BLOOD:   |                 |                     |                      |
| RX     | MEDLINE: 96369309.  |                 |                     |                      |
| RA     | LIN L., TOKUNAGA K., TANAKA H., NAKAJIMA F., IMANISHI T.,     |                 |                     |                      |
| RA     | JUJI T.,  |                 |                     |                      |
| RL     | TISSUE ANTIGENS 47:265-274(1996).                             |                 |                     |                      |
| EMBL   | D44500; G1552171; -.  |                 |                     |                      |
| DR     | PROSITE: PS00290; IG_MHC; 1.                                  |                 |                     |                      |
| KW     | MHC.  |                 |                     |                      |
| FT     | NON_TER   | 298             |                     |                      |
| SO     | SEQUENCE  | 298 AA;         | 34222 MW;           | 31AE7BCE CRC32;      |
|        | Query Match   | 80.0%;          | Score 64;           | DB 5; Length 298;    |
|        | Best Local Similarity   | 88.9%;          | Pred. No. 6,77e-01; |                      |
|        | Matches   | 8; Conservative | 0; Mismatches       | 1; Indels 0; Gaps 0; |
| Db     | 233 YPAEITLTW 241   |                 |                     |                      |
| QY     | 1 YPAEITLTW 9   |                 |                     |                      |

```

RESULT 7
ID 019652 PRELIMINARY; PRT; 298 AA.
AC 019652;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HLA-CW*1203 (FRAGMENT).
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PERIPHERAL BLOOD;
RA WANG H., TOKUNAGA K.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; D64146; D1020302; -.
DR PROSITE; PS00290; IG_MHC; 1.
KW MHC.
FT NON_TER 298 298
SQ SEQUENCE 298 AA; 34375 MW; CCL1A67C0 CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 5; Length 298;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 233 YPAEITLW 241
1 YPAEITLW 9

RESULT 8
ID 095349 PRELIMINARY; PRT; 298 AA.
AC 095349;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HLA-B*1528 (FRAGMENT).
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD;
RX MEDLINE; 96369309.
RA LIN L., TOKUNAGA K., TANAKA H., NAKAJIMA F., IMANISHI T.,
RA KASHIMASE K., BANNAI M., MIZUNO S., ARAZA T., TADOKORO K., SHIBATA Y.,
RA JUJI T.;
RL TISSUE ANTIGENS 47:265-274(1996).
DR EMBL; D44489; G1552170; -.
DR PROSITE; PS00290; IG_MHC; 1.
KW MHC.
FT NON_TER 298 298
SQ SEQUENCE 298 AA; 34268 MW; B2AF2342 CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 5; Length 298;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 233 YPAEITLW 241
1 YPAEITLW 9

RESULT 9
ID 019653 PRELIMINARY; PRT; 298 AA.
AC 019653;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HLA-CW*0304 (FRAGMENT).
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

```

```

OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PERIPHERAL BLOOD;
RA WANG H., TOKUNAGA K.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; D64150; D1020306; -.
DR PROSITE; PS00290; IG_MHC; 1.
KW MHC.
FT NON_TER 298 298
SQ SEQUENCE 298 AA; 34352 MW; A298DB06 CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 5; Length 298;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 233 YPAEITLW 241
1 YPAEITLW 9

RESULT 10
ID 029866 PRELIMINARY; PRT; 298 AA.
AC 029866;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HLA CLASS I HLA-C ALLELE PRECURSOR (FRAGMENT).
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD;
RA MARGRET M., BROCKSTEEDT D., JENISCH S.;
RL SUBMITTED (MAY-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; 233459; G488362; -.
DR PROSITE; PS00290; IG_MHC; 1.
KW SIGNAL; MHC.
FT SIGNAL 1 24
FT CHAIN 25 >298
FT NON_TER 298 298
SQ SEQUENCE 298 AA; 34312 MW; F45DABE CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 5; Length 298;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 233 YPAEITLW 241
1 YPAEITLW 9

RESULT 11
ID 019655 PRELIMINARY; PRT; 298 AA.
AC 019655;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HLA-CW*0801 (FRAGMENT).
OS HOMO SAPIENS (HUMAN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PERIPHERAL BLOOD;
RA WANG H., TOKUNAGA K.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; D64151; D1020307; -.
DR PROSITE; PS00290; IG_MHC; 1.
KW MHC.
FT NON_TER 298 298
SQ SEQUENCE 298 AA; 34304 MW; 84207277 CRC32;

```

Query Match 80.0%; Score 64; DB 5; Length 298;  
Best Local Similarity 88.9%; Pred. No. 6.77e-01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241  
1 YPAEITLW 9

RESULT 12  
ID 095392 PRELIMINARY; PRT; 302 AA.

AC 095392;  
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE LEUKOCYTE ANTIGEN B (HLA-B\*47 VARIANT) (FRAGMENT).

GN HLA-B.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.

RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PERIPHERAL BLOOD;  
RA FISCHER G.F., FAE I., LEITNER D., MAYR W.R.;  
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Y09118; E280547; -;  
DR PROSITE; PS00290; IG\_MHC; 1.

KW MHC.  
FT NON\_TER 302 302  
SQ SEQUENCE 302 AA; 34828 MW; D395D4B7 CRC32;

Query Match 80.0%; Score 64; DB 5; Length 302;  
Best Local Similarity 88.9%; Pred. No. 6.77e-01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241  
1 YPAEITLW 9

RESULT 13  
ID 029654 PRELIMINARY; PRT; 322 AA.

AC 029654;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE HLA-B\*71 (FRAGMENT).  
GN B-1510.

OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.

RN [1]  
RP SEQUENCE FROM N.A.  
RA HORLEY C.K., BEI M., RODRIGUEZ S., JOHNSON A.;  
RL SUBMITTED (JUL-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U11264; G511780; -;  
DR PROSITE; PS00290; IG\_MHC; 1.

KW MHC.  
FT NON\_TER 322 322  
SQ SEQUENCE 322 AA; 36634 MW; 7D650C0C CRC32;

Query Match 80.0%; Score 64; DB 5; Length 322;  
Best Local Similarity 88.9%; Pred. No. 6.77e-01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241  
1 YPAEITLW 9

RESULT 14  
ID 030207 PRELIMINARY; PRT; 338 AA.

AC 030207;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE HLA-B\*62 ANTIGEN (FRAGMENT).  
GN HLA-B\*62.4.

OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.

RN [1]  
RP SEQUENCE FROM N.A.  
RA CHOO S.Y., FAN L.A., HANSEN J.A.;  
RL SUBMITTED (AUG-1992) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: M83194; G292920; -;  
DR PROSITE; PS00290; IG\_MHC; 1.

KW MHC.  
FT NON\_TER 1 1  
SQ SEQUENCE 338 AA; 37759 MW; A6054B6D CRC32;

Query Match 80.0%; Score 64; DB 5; Length 338;  
Best Local Similarity 88.9%; Pred. No. 6.77e-01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 209 YPAEITLW 217  
1 YPAEITLW 9

RESULT 15  
ID 030443 PRELIMINARY; PRT; 356 AA.

AC 030443;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE MHC CLASS I CAJA-G\*03 (FRAGMENT).

GN CAJA-G.  
OS CALLITHRIX JACCHUS (COMMON MARMOSET).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.

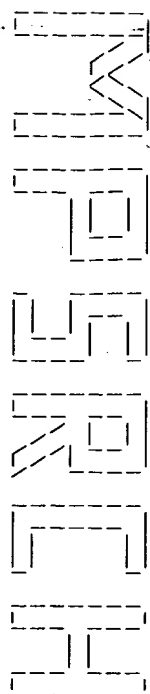
RN [1]  
RP SEQUENCE FROM N.A.  
RA CADAVID L.F., SHUEFLEBOTHAM C., RUIZ F.J., YEAGER M., HUGHES A.L.;  
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U59639; G1389925; -;

KW MHC.  
FT NON\_TER 1 1  
SQ SEQUENCE 356 AA; 39987 MW; 9528E499 CRC32;

Query Match 80.0%; Score 64; DB 5; Length 356;  
Best Local Similarity 88.9%; Pred. No. 6.77e-01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 225 YPAEITLW 233  
1 YPAEITLW 9

Search completed: Fri Sep 11 12:51:52 1998  
Job time : 31 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch.p: protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 12:45:28 1998; MasPar time 2.62 Seconds

Tabular output not generated. 67.915 Million cell updates/sec

Title: >US-08-452-843-3  
Description: (1-11) from US08452843.pep  
Perfect Score: 89  
Sequence: 1 CITESECFRAVI 11

Scoring table:  
PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 17.702; Variance 53.911; scale 0.328

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID  | Description            | Pred. No. |
|------------|-------|-------------|--------|--------|------------------------|-----------|
| 1          | 89    | 100.0       | 11 18  | R89364 | MAGE-1 derived immuno  | 4.13e-03  |
| 2          | 67    | 75.3        | 309 13 | R70909 | Human melanoma antige  | 1.53e+00  |
| 3          | 58    | 65.2        | 615 6  | R31349 | Jaagsiekte retrovirus  | 1.53e+01  |
| 4          | 56    | 62.9        | 477 25 | W25153 | Nsp7524111 restrictio  | 2.52e+01  |
| 5          | 53    | 59.6        | 9 15   | R78910 | MAGE 1 92-100 cytocto  | 5.26e+01  |
| 6          | 53    | 59.6        | 9 12   | R73821 | Antigen fragment 137,  | 5.26e+01  |
| 7          | 53    | 59.6        | 200 17 | R94701 | PRSV VR 2385 ORF-5 p   | 5.26e+01  |
| 8          | 53    | 59.6        | 200 24 | W25953 | ORF 5 protein of PRS   | 5.26e+01  |
| 9          | 53    | 59.6        | 455 5  | R28757 | Hepatocyte nuclear fa  | 5.26e+01  |
| 10         | 52    | 58.4        | 489 9  | R44295 | Corticosterone-bindin  | 6.70e+01  |
| 11         | 50    | 56.2        | 308 16 | R87723 | Full length cocoonut l | 1.08e+02  |
| 12         | 50    | 56.2        | 358 22 | W20718 | H. pylori membrane pr  | 1.08e+02  |
| 13         | 50    | 56.2        | 591 14 | R74802 | Saccharomyces sp. rec  | 1.08e+02  |
| 14         | 49    | 55.1        | 251 29 | W47421 | Bacillus subtilis pre  | 1.38e+02  |
| 15         | 49    | 55.1        | 286 22 | W24585 | H. pylori cytoplasmic  | 1.38e+02  |
| 16         | 49    | 55.1        | 286 22 | W20102 | H. pylori cytoplasmic  | 1.38e+02  |
| 17         | 49    | 55.1        | 455 22 | W20606 | H. pylori cytoplasmic  | 1.38e+02  |
| 18         | 49    | 55.1        | 500 25 | W30843 | Partial rat thrombomo  | 1.38e+02  |

|    |    |      |         |        |                        |          |
|----|----|------|---------|--------|------------------------|----------|
| 19 | 49 | 55.1 | 559 25  | W30844 | Partial rat thrombomo  | 1.38e+02 |
| 20 | 49 | 55.1 | 577 25  | W30845 | Rat thrombomodulin.    | 1.38e+02 |
| 21 | 49 | 55.1 | 889 10  | R56248 | Xenopus thrombospondi  | 1.38e+02 |
| 22 | 49 | 55.1 | 3164 16 | R94345 | Hepatitis GB virus (H  | 1.38e+02 |
| 23 | 48 | 53.9 | 55 12   | R63534 | HT-ICP fragment, cor   | 1.74e+02 |
| 24 | 48 | 53.9 | 94 12   | R63532 | Leukocyte Chemotactic  | 1.74e+02 |
| 25 | 48 | 53.9 | 105 12  | R63531 | Leukocyte Chemotactic  | 1.74e+02 |
| 26 | 48 | 53.9 | 106 12  | R63533 | Recombinant (Met)-HT-  | 1.74e+02 |
| 27 | 48 | 53.9 | 337 29  | W40137 | Human partial GAI2 r   | 1.74e+02 |
| 28 | 48 | 53.9 | 387 25  | W24562 | Human galactin recepto | 1.74e+02 |
| 29 | 48 | 53.9 | 580 17  | R93607 | Kaposi's sarcoma asso  | 1.74e+02 |
| 30 | 48 | 53.9 | 580 18  | R97831 | Kaposi's sarcoma asso  | 1.74e+02 |
| 31 | 48 | 53.9 | 896 12  | R63533 | Human HT-1376 cell-de  | 1.74e+02 |
| 32 | 47 | 52.8 | 9 11    | R59206 | Peptide fragment (1.0  | 2.20e+02 |
| 33 | 47 | 52.8 | 349 27  | W35849 | Human CD5 for use in   | 2.20e+02 |
| 34 | 47 | 52.8 | 397 14  | R75681 | Human Wnt-x.           | 2.20e+02 |
| 35 | 47 | 52.8 | 612 28  | W39256 | Human partial mature   | 2.20e+02 |
| 36 | 47 | 52.8 | 737 28  | W39257 | Human membrane protei  | 2.20e+02 |
| 37 | 47 | 52.8 | 961 10  | R56249 | Human thrombospondin   | 2.20e+02 |
| 38 | 46 | 51.7 | 184 19  | R98994 | Vascular IBP-like gro  | 2.78e+02 |
| 39 | 46 | 51.7 | 190 19  | R91710 | AcenAP4.               | 2.78e+02 |
| 40 | 46 | 51.7 | 672 17  | R94765 | Type III (alpha-type)  | 2.78e+02 |
| 41 | 46 | 51.7 | 672 13  | R66726 | Protein kinase C muta  | 2.78e+02 |
| 42 | 46 | 51.7 | 790 18  | R98900 | Squash mosaic virus s  | 2.78e+02 |
| 43 | 46 | 51.7 | 874 20  | W06358 | E6-associated protein  | 2.78e+02 |
| 44 | 46 | 51.7 | 874 14  | R79657 | Human E6-AP protein.   | 2.78e+02 |
| 45 | 46 | 51.7 | 1074 4  | R24102 | Marek's disease virus  | 2.78e+02 |

# ALIGNMENTS

RESULT 1  
ID R89364 standard; peptide: 11 AA.  
AC R89364;  
DT 18-SEP-1996 (first entry)  
DE MAGE-1 derived immunogenic peptide.  
KW immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI, 96-116784/12.  
PT Compos. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g. the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 11 AA;

Query Match 100.0%; Score 89; DB 18; Length 11;  
Best Local Similarity 100.0%; Pred. No. 4.13e-03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 CITESECFRAVI 11  
QY 1 CITESECFRAVI 11

RESULT 2  
ID R70909 standard; Protein: 309 AA.  
AC R70909;  
DT 09-OCT-1995 (first entry)  
DE Human melanoma antigen; MAGE-1.  
KW Human melanoma antigen; MAGE-1; vaccines; MAGE associated tumours;  
KW HLA-restricted cytotoxic T-lymphocyte activity.  
OS Homo sapiens.  
PN WO9504542-A.  
PD 16-FEB-1995.  
PF 02-AUG-1994; U08721.  
PR 06-AUG-1993; US-103623.  
PA (CYTE-) CYTEL CORP.  
PI Files JD, Livingston BD, Sette AD, Sidney JC;  
DR WPI: 95-090681/12.  
N-PSDB: Q85435.  
PT Human melanoma antigen, MAGE-1, peptide(s) - useful for  
stimulating immune response against melanoma  
PS Example 1: Fig 1: 59pp; English.  
CC 085435 encodes R70909 human melanoma antigen MAGE-1, it was used  
to produce the C-terminal MAGE-1 peptides described in R70915 to  
R70969. These peptides are useful for defining epitopes that  
engender a HLA-restricted cytotoxic lymphocyte activity against  
MAGE-1 antigens. Compsns. containing these peptides can be  
administered, as a vaccine to patients susceptible to MAGE  
associated tumours, e.g. melanomas.  
SQ Sequence 309 AA;

Query Match 75.3%; Score 67; DB 13; Length 309;  
Best Local Similarity 90.9%; Pred. No. 1.53e+00;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 92 cilesfravi 102  
|||||  
1 CILESCFRAY 11

RESULT 3  
ID R31349 standard; Protein: 615 AA.  
AC R31349;  
DT 18-MAY-1993 (first entry)  
DE Jaagsiekte retrovirus Env protein.  
KW JSRV; epithelial carcinoma; ovine; sheep; vaccine;  
KW pulmonary adenomatosa; envelope glycoprotein.  
OS Jaagsiekte retrovirus.  
FH Key Location/Qualifiers  
FT region 1..378  
FT /note="surface portion"  
FT 379..615  
FT /note="transmembrane portion"  
FT region  
FR2676455-A.  
PD 20-NOV-1992.  
PF 17-MAY-1991; 006060.  
PR 17-MAY-1991; FR-006060.  
PA (INRM) INSERM INST NAT SANTE & RECH MED.  
PI Querat GF, Verwoerd D, Vigne R, York D;  
DR WPI: 93-020250/03.  
N-PSDB: Q35153.  
DE New Jaagsiekte Retrovirus and characteristic nucleic acid - also  
PT derived proteins, probes and antibodies, useful for in vitro  
diagnosis and in vaccines  
PS Claim 26; Page 41-43; 75pp; French.  
CC JSRV causes epithelial carcinoma in ovine animals, partic. pulmonary  
adenomatosa in sheep. The complete cDNA sequence of the JSRV genome  
was prepared from an approx. 8.7kb band of poly-A RNA isolated from  
semi-purified lung lavage samples from infected sheep. The  
invention includes the Env amino acid sequence or any part of it  
which is capable of specific immunological reaction with antibodies  
directed against JSRV. The glycoproteins gp46 or gp31 and the  
precursor PR69 env are preferred.  
CC See also R31346-R31348 and Q35153-Q35155.  
SQ Sequence 615 AA;

Query Match 65.2%; Score 58; DB 6; Length 615;  
Best Local Similarity 60.0%; Pred. No. 1.53e+01;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 323 ciltncirgv 332  
|||:|:|:  
1 CILESCFRAY 10

RESULT 4  
ID W25153 standard; Protein: 477 AA.  
AC W25153;  
DT 22-JAN-1998 (updated)  
DE 03-DEC-1997 (first entry)  
DE Nsp524IIR restriction enzyme isoform.  
KW Restriction enzyme; NspIII; Nostoc species; genetic engineering;  
KW cloning; vector construction; recombinant production; endonuclease.  
OS Nostoc sp.  
PN J09191885-A.  
PD 29-JUL-1997.  
PF 16-JAN-1996; 023304.  
PR 16-JAN-1996; JP-023304.  
PA (TAKI) TAKARA SHUZO CO LTD.  
DR WPI: 97-429185/40.  
N-PSDB: T79876, T89627.  
PT Nsp524IIR restriction endonuclease and its gene - useful in  
genetic engineering methods, e.g. vector construction and cloning  
PS Claim 9; Page 11-12; 15pp; Japanese.  
CC W25153 shows the sequence of an Nsp524IIR restriction enzyme isoform  
(derived from Nostoc sp. PCC7524). The full length gene encoding this  
enzyme also contains a second open reading frame encoding a similar  
but different NspIII enzyme.  
CC NspIII restriction endonucleases are useful in genetic engineering  
methods such as vector construction and cloning. The enzymes cut  
between the first and second nucleotides of the sequence CYGRC.  
CC Nsp524IIR can be produced recombinantly in a large amount.  
CC (Revised entry submitted to correct crossreference to N-PSDB.)  
SQ Sequence 477 AA;

Query Match 62.9%; Score 56; DB 25; Length 477;  
Best Local Similarity 40.0%; Pred. No. 2.52e+01;  
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 396 cvlgecfv1 405  
||:|:|:  
1 CILESCFRAY 10

RESULT 5  
ID R78910 standard; peptide: 9 AA.  
AC R78910;  
DT 27-MAR-1996 (first entry)  
DE MAGE 1 92-100 cytotoxic T lymphocyte epitope.  
KW MAGE 1 92-100: cytotoxic T; CTL; epitope; helper T; HTL; cell;  
KW lymphocyte; antigens; treatment; disease prevention; tumours;  
KW cancer; melanomas.  
OS Homo sapiens.  
PN WO9522317-A1.  
PD 24-AUG-1995.  
PR 16-FEB-1995; U02121.  
PA 16-FEB-1994; US-197484.  
CYTE-) CYTEL CORP.  
PI Ceut RW, Grey H, Sette AD, Vitelli MA;  
DR WPI: 95-302545/39.  
PT Compns. inducing cytotoxic T lymphocyte response to pref. viral,  
bacterial, parasitic or tumour antigens - useful in the treatment  
PT and prevention of diseases associated with the antigen e.g.  
hepatitis B  
PS Example 13; Page 71; 109pp; English.  
CC A compsn. which induces a cytotoxic T lymphocyte (CTL) response to  
a human MAGE antigen (Ag) in a mammal comprises, a MAGE CTL Ag  
response inducing peptide (i.e. R78904 to R78917) and a lipid

CC conjugated helper T cell inducing peptide. The compsn. is useful  
 CC in the treatment and prevention of MAGE tumour Ag associated  
 CC diseases, e.g. melanoma cancers.  
 SO Sequence 9 AA;

Query Match 59.6%; Score 53; DB 15; Length 9;  
 Best Local Similarity 88.9%; Pred. No. 5.26e+01;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 cilescifra 9  
 ||||| |||  
 QY 1 CILESCFRA 9

RESULT 6  
 ID R73821 standard; peptide: 9 AA.

AC R73821.  
 DE 22-JUN-1995 (first entry)  
 DT Antigen fragment 137, from MAGE1 has binding affinity for HLA-A2.1.  
 KW antigen; epitope; immunogenic target protein; PSA; HBVC; HBVs; EBV;  
 KW HIV1; plasma specific antigen; hepatitis B virus; Epstein Barr;  
 KW human immunodeficiency virus; human papilloma virus; P53; c-erbB2;  
 KW MAGE-1; melanoma antigen-1; core antigen; suicide antigen;  
 KW pharmaceutical composition; in vivo; ex vivo; therapeutic;  
 KW diagnostic; MHC class I molecule; major histocompatibility complex;  
 KW HLA-A2.1; 9mer; 10mer; anchor; human leukocyte antigen; PLP; 8mer;  
 KW algorithm prediction; MBP; CMV; cytomegalovirus; HSV;  
 KW herpes simplex virus; influenza A; ML; LCMV.  
 OS Homo sapiens.  
 PN W09420127-A.  
 PD 15-SEP-1994.  
 PE 04-MAR-1994; US-020253.  
 PR 05-MAR-1993; US-027146.  
 PR 04-JUN-1993; US-073205.  
 PR 29-NOV-1993; US-159184.  
 PA (CYTE-) CYTEL CORP.  
 PI Grey HM, Kast WM, Sette A, Sidney J;  
 DR WPI: 94-302678/37.

PT Immunogenic peptide(s) having an HLA-A2.1 binding motif - used  
 PT for treatment or prophylaxis of cancer, virus infection or  
 PT autoimmune diseases.  
 PS Disclosure: Page 85; 138pp: English.  
 PS R73685-876 are potential peptide binders of HLA-A2.1 motif. Using  
 CC motifs disclosed in the invention, these peptides were screened for  
 CC further motifs. Only peptides with binding affinity of at least 18  
 CC (binding affinity is expressed as an IC50 value) as compared to the  
 CC standard peptide (R71293) in assays. This peptide from MAGE1 has a  
 CC binding value of 0.0460. The peptides of the invention can induce  
 CC cytotoxic T lymphocytes which can react with target cells. They can  
 CC be used for the treatment or prophylaxis of cancer, eg. prostate  
 CC cancer or lymphoma, etc.  
 SO Sequence 9 AA;

Query Match 59.6%; Score 53; DB 12; Length 9;  
 Best Local Similarity 88.9%; Pred. No. 5.26e+01;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 cilescifra 9  
 ||||| |||  
 QY 1 CILESCFRA 9

RESULT 7  
 ID R94701 standard; Protein; 200 AA.

AC R94701.  
 DE 04-AUG-1996 (first entry)  
 DT PRRSV VR 2385 ORF-5 product.  
 KW PRRSV; vaccine; antigen.  
 OS Pig reproductive and respiratory syndrome virus Iowa strain ISU-12.  
 PN W09606619-A1.  
 PD 07-MAR-1996.  
 PE 01-SEP-1995; U10904.  
 PR 01-SEP-1994; US-301435.

PA (HALB/) HALBUR P.  
 PA (LUMM/) LUM M. A.  
 PA (MENG/) MENG X.  
 PA (MORO/) MOROZOV I.  
 PA (PAUL/) PAUL P. S.  
 PI Halbur P, Lum MA,  
 DR WPI: 96-160132/16.  
 DR N-PSDB: T14390.

PT New porcine reproductive and respiratory syndrome virus DNA - and  
 PT proteins encoded by open reading frames of an Iowa strain of the  
 PT virus; are used in vaccines against PRRSV in pigs  
 PS Disclosure: Page 140-141; 228pp: English.  
 CC The protein (R94701) encoded by open reading frame 5 (ORF-5 -  
 CC T14390) of porcine reproductive and respiratory syndrome virus  
 CC (PRRSV) Iowa strain isolate ISU-12 (VR 2385) can be used with  
 CC other PRRSV proteins (see also R94702-03, R94707-16 and  
 CC R94719-21) in the development of vaccines against PRRSV in pigs  
 CC and in serological tests for screening pigs for exposure to, or  
 CC infection by, PRRSV (partic. strain Iowa).  
 SO Sequence 200 AA;

Query Match 59.6%; Score 53; DB 17; Length 200;  
 Best Local Similarity 54.5%; Pred. No. 5.26e+01;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 19 clyscfvaly 29  
 ||: ||| ||:  
 QY 1 CILESCFRAV1 11

RESULT 8  
 ID W25953 standard; Protein; 200 AA.  
 AC W25953.  
 DE 10-NOV-1997 (first entry)  
 DT ORF 5 protein of PRRSV isolate VR2385.  
 KW Porcine reproductive and respiratory syndrome virus; PRRSV; coronavirus;  
 KW reproductive failure; pneumonia; pig; preweaning mortality; torovirus;  
 KW subgenomic mRNA; glycosylated membrane protein; nucleocapsid protein;  
 KW membrane associated protein; vaccine; antibody; therapy.  
 OS Porcine reproductive and respiratory syndrome virus.  
 PN W09640932-A1.  
 PD 19-DEC-1996.  
 PE 07-JUN-1996; U08962.  
 PR 07-JUN-1995; US-478316.  
 PA (HALB/) HALBUR P.  
 PA (MENG/) MENG X.  
 PA (MORO/) MOROZOV I.  
 PA (PAUL/) PAUL P. S.  
 PI Halbur P, Meng X, Morozov I, Paul PS;  
 DR WPI: 97-108646/10.  
 DR N-PSDB: T60795.

PT Porcine reproductive and respiratory syndrome virus DNA sequences -  
 PT useful for diagnosis, treatment and prevention of infection in pigs  
 PS Disclosure: Fig 2d; 114pp: English.  
 CC W25950-W25977 represent proteins encoded by ORFs 2-5 of different  
 CC isolates of porcine reproductive and respiratory syndrome virus (PRRSV).  
 CC PRRSV is a new and severe disease in swine, characterised by reproductive  
 CC failure in sows and guinea pigs, pneumonia in young growing pigs, and an  
 CC increase in preweaning mortality. However, there are marked differences in  
 CC pathogenicity between isolates (with ISU3927 being the least virulent  
 CC isolate known). The genomic organisation of PRRSV resembles coronaviruses  
 CC and toroviruses, in that their replication involves the formation of a  
 CC 3'-coterminal nested set of subgenomic mRNAs. ORFs 5, 6, and 7 encode a  
 CC glycosylated membrane protein, an unglycosylated membrane protein, and a  
 CC nucleocapsid protein, respectively. ORFs 2 to 4 encode proteins with the  
 CC characteristics of membrane associated proteins. The polynucleotides of  
 CC the invention, encode a protein that is at least 88%, but less than 100%  
 CC homologous to one of proteins encoded by one of the ORFs of these  
 CC sequences. The polynucleotides of the invention, and their encoded  
 CC polypeptides can be used in a vaccine to protect a pig against PRRSV.  
 CC Antibodies raised against the polypeptides can be used to treat a pig  
 CC suffering from PRRSV, and to assay for a PRRSV.  
 SO Sequence 200 AA;





Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 15 lescfka 21

OY 3 LESCFA 9

# RESULT 12

ID W20718 standard; protein; 358 AA.

AC W20718; 13-JUL-1997 (first entry)  
DE H. pylori membrane protein, 05cp20518orf61.  
KW Vaccine; prevention; treatment; infection; identification;  
binding compound; bacterium; life cycle; activator; bacteria;  
inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
KW membrane; amino acid; metabolism.  
OS Helicobacter pylori.  
PN W09640893-A1.  
PD 19-DEC-1996.  
PF 06-JUN-1996; U09122.  
PR 07-JUN-1995; US-487032.  
PR 01-APR-1996; US-630405.  
PA (ASTR) ASTRA AB.  
PI Bergholm OT, Smith D, Mellgaard BL;  
DR WPI; 97-052306/05.  
PT Helicobacter pylori nucleic acid sequences and related  
polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
infection, and to detect Helicobacter  
PS Claim 73; Pages 1133-1134; 1481pp; English.  
CC The present sequence is a Helicobacter pylori membrane protein  
CC likely to contain four membrane spanning regions.  
CC The protein may be used in a vaccine to prevent or treat  
H. pylori infection or to identify H. pylori polypeptide binding  
CC compounds, useful as potential H. pylori life cycle activators or  
CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was  
CC determined from overlapping contigs generated by mechanically  
CC shearing the bacterial DNA. The sequences were analysed for ORF of  
CC at least 180 nucleotides, and the predicted coding regions defined  
CC by computer evaluation. To identify likely H. pylori antigens for  
CC vaccine development, the amino acid sequences predicted from  
CC various ORF were analysed for significant homology to other known  
CC or exported membrane proteins. Having identified and determined  
CC the sequences of interest, particular regions can be isolated from  
CC H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
SQ Sequence 358 AA;

Query Match 56.2%; Score 50; DB 22; Length 358;

Best Local Similarity 50.0%; Pred. No. 1.08e+02;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 250 cs1gcscfcal 259

OY 1 CILESCFRAV 10

# RESULT 13

ID R74802 standard; Protein; 591 AA.

AC R74802; 21-NOV-1995 (first entry)  
DE Saccharomyces sp. recombinant xylokinase.  
KW Xylokinase; enzyme; xylose; ethanol; fuel; yeast; fungus.  
OS Saccharomyces sp.  
PN W09513362-A.  
PD 18-MAY-1995.  
PF 08-NOV-1994; U12861.  
PR 08-NOV-1993; US-148581.  
PA (PURD) PURDUE RES FOUND.  
PI Ho NWY Tsao GT;  
DR WPI; 95-194082/25.  
DR N-PSDB; Q90271.  
PT Recombinant yeast encoding xylose reductase, xyitol dehydrogenase and  
PT xylokinase - can effectively ferment xylose alone, or

PT simultaneously with glucose, to produce ethanol e.g for use as a fuel  
PS disclosure; Fig.2; 63pp; English.

CC This sequence encoding xylokinase (XK) is used along with those  
CC encoding xylose-reductase and xyitol-dehydrogenase to effect  
CC fermentation of xylose to fuel ethanol when expressed by recombinant  
CC Saccharomyces sp. XK is able to convert D-xylose to D-xylose-  
CC 5-phosphate.  
SQ Sequence 591 AA;

Query Match 56.2%; Score 50; DB 14; Length 591;

Best Local Similarity 40.0%; Pred. No. 1.08e+02;

Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 535 calgcyckem 544

OY 1 CILESCFRAV 10

# RESULT 14

ID W47421 standard; Protein; 251 AA.

AC W47421; 05-JUN-1998 (first entry)  
DE Bacillus subtilis prenyl diphosphate synthetase subunit.  
KW Prenyl diphosphate synthetase; subunit; polyprenyl diphosphate;  
vitamin K; ubiquinone.  
OS Bacillus subtilis.  
PN EP-812914-A2.  
PD 17-DEC-1997.  
PF 13-JUN-1997; 109692.  
PR 14-JUN-1996; JP-154441.  
PA (TOYT) TOYOTA JIDOSHA KK.  
PI Cho Y, Kolke A, Koyama T, Muramatsu M, Ogura K,  
PI Shimizu N;  
DR WPI; 98-034975/04.  
PT DNA encoding prenyl diphosphate synthetase subunit(s) - new  
PT Micrococcus prenyl diphosphate synthetase subunit polypeptide(s),  
PS and methods for preparing enzymes from subunit(s)  
PS Claim 14; Pages 23-24; 46pp; English.  
CC A thermally resistant prenyl diphosphate synthetase (PDS) comprises  
CC a PDS subunit from Bacillus subtilis (W47421), and a PDS subunit  
CC from B. stearothermophilus (W47422).  
CC Substances synthesised by PDS, i.e. polyprenyl diphosphates, are  
CC precursors of physiologically active substances, e.g. vitamin K and  
CC ubiquinones.  
SQ Sequence 251 AA;

Query Match 55.1%; Score 49; DB 29; Length 251;

Best Local Similarity 71.4%; Pred. No. 1.38e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 205 l1edcfk 211

OY 2 ILESCFR 8

# RESULT 15

ID W24585 standard; Protein; 286 AA.

AC W24585; 04-AUG-1997 (first entry)  
DE H. pylori cytoplasmic protein, 11253. aa.  
KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;  
secreted; periplasmic; chronic gastritis; duodenal ulcer disease;  
KW activator; inhibitor; bacterial life cycle; vaccine; immunise;  
KW detection; antisense; inhibition.  
OS Helicobacter pylori.  
FH Key location/Qualifiers  
FT misc\_difference 2 /note= "encoded by AAS"  
FT misc\_difference 133 /note= "encoded by TYA"  
FT misc\_difference 166 /note= "encoded by RCC"  
FT W09719098-A1.

PD 29-MAY-1997. 15-NOV-1996; U18542.  
PF 17-NOV-1995; US-561469.  
PA (ASTR ) ASTRAB.  
PI Smith DH;  
DR WPI; 97-298052/27.  
DR N-PSDB; T77043.  
PT Helicobacter pylori nucleic acid sequences and related proteins -  
PS used for diagnostics and therapeutics  
PS Claim 18; Page 131; 235p; English.  
CC This sequence represents an H. pylori cytoplasmic protein involved in  
CC outer membrane or cell wall biosynthesis. This sequence showed  
CC homology to N-acetylmuramate-Alanine ligase.  
CC Helicobacter pylori has been strongly linked to chronic gastritis and  
CC duodenal ulcer disease. The nucleic acid sequences of the invention  
CC are used to evaluate compounds, especially activators or inhibitors of  
CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid  
CC sequence. The nucleic acid sequences, and corresponding proteins, are  
CC also useful for generating vaccines for immunising subjects against H.  
CC pylori or for use in detecting the presence of Helicobacter species in  
CC a sample. Antisense nucleic acid sequences of these sequences are  
CC used to inhibit expression of a gene from Helicobacter species. H.  
CC pylori whole genomic DNA was isolated and nebulised to a median size of  
CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique  
CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are  
CC complementary to the BstXI-cut PMPX vectors, while the overhang is not  
CC self-complementary. Therefore the linkers will not concatamerise nor  
CC will the cut vector re-ligate itself easily. The linker-adaptor inserts  
CC were ligated to each of the 20 PMPX vectors to construct a series of  
CC shotgun subclone libraries. The purified DNA samples were then  
CC sequenced.  
CC Note: The ORF/protein reference number for this sequence was obtained  
CC from the related specification, WO9640893.  
SQ Sequence 286 AA;

Query Match 55.1%; Score 49; DB 22; Length 286;  
Best Local Similarity 54.5%; Pred. No. 1.38e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
Db 206 ctlehcdrl11 216  
QY 1 CILESCRAVI 11

Search completed: Fri Sep 11 12:45:44 1998  
Job time : 16 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

Mpsrch_pp 'protein' - protein database search, using Smith-Waterman algorithm
Run on:      Fri Sep 11 12:46:03 1998;  MasPar time 3.44 Seconds
Tabular output not generated.             110.713 Million cell updates/sec

```

|                |                            |
|----------------|----------------------------|
| Title:         | >US-08-452-843-3           |
| Description:   | (1-11) from US08452843.pep |
| Perfect Score: | 89                         |
| Sequence:      | 1 CILESCRAVI 11            |

Scoring table: PAM 150  
Gap 15

Searched: 120441 seqs, 3653193 residues

post-processing: Minimum Match 08  
Listing first 45 summaries

```
Database: r
pir56
1:pir1 2:pir2 3:pir3 4:pir4 5:nr13d
```

Statistics: Mean 23.859; Variance 36.336; scale 0.657

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID      | Description           | Pred. No. |
|------------|-------|-------------|--------|----|---------|-----------------------|-----------|
| 1          | 67    | 75.3        | 280    | 2  | JC3358  | tumor-associated anti | 3.35e-020 |
| 2          | 61    | 68.5        | 1583   | 2  | S55644  | hypothetical protein  | 4.56e-01  |
| 3          | 58    | 65.2        | 615    | 1  | VCWVJA  | env polyprotein precu | 1.60e+00  |
| 4          | 58    | 65.2        | 2118   | 2  | J272705 | myceroctic acid synt  | 1.60e+00  |
| 5          | 57    | 62.9        | 3175   | 1  | RRWREV  | RNA-directed RNA pol  | 2.42e+00  |
| 6          | 56    | 60.0        | 427    | 1  | QOEC03  | hypothetical 49.6 kd  | 3.63e+00  |
| 7          | 54    | 60.7        | 130    | 2  | A33893  | foliitroph beta chlt  | 8.10e+00  |
| 8          | 54    | 60.7        | 647    | 2  | C68102  | DNA mismatch recognit | 8.10e+00  |
| 9          | 53    | 59.6        | 330    | 2  | I55975  | X/Y protein - mouse ( | 1.20e+01  |
| 10         | 53    | 59.6        | 408    | 2  | JC4938  | hepatocyte nuclear fa | 1.20e+01  |
| 11         | 53    | 59.6        | 433    | 2  | A30550  | complement C3b/C4b re | 1.20e+01  |
| 12         | 53    | 59.6        | 438    | 2  | S55631  | virion protein kinase | 1.20e+01  |
| 13         | 53    | 59.6        | 440    | 2  | A43519  | complement receptor C | 1.20e+01  |
| 14         | 53    | 59.6        | 455    | 2  | JC4936  | hepatocyte nuclear fa | 1.20e+01  |
| 15         | 53    | 59.6        | 455    | 2  | A36471  | transcription factor  | 1.20e+01  |
| 16         | 53    | 59.6        | 463    | 2  | JC4009  | hepatocyte nuclear fa | 1.20e+01  |
| 17         | 53    | 59.6        | 465    | 2  | S25074  | hepatocyte nuclear fa | 1.20e+01  |
| 18         | 53    | 59.6        | 465    | 2  | S52052  | hepatocyte nuclear fa | 1.20e+01  |
| 19         | 53    | 59.6        | 465    | 2  | JC4937  | hepatocyte nuclear fa | 1.20e+01  |
| 20         | 53    | 59.6        | 504    | 2  | JC6096  | hepatocyte nuclear fa | 1.20e+01  |
| 21         | 53    | 59.6        | 1377   | 2  | A38926  | DNA-binding protein c | 1.20e+01  |
| 22         | 52    | 58.4        | 117    | 2  | C37330  | venom allergen IV - r | 1.77e+01  |
| 23         | 52    | 58.4        | 127    | 2  | D65186  | hypothetical protein  | 1.77e+01  |

|    |    |      |      |   |        |                       |          |
|----|----|------|------|---|--------|-----------------------|----------|
| 25 | 52 | 58.4 | 129  | 2 | A48333 | probable membrane pro | 1.77e+01 |
| 24 | 52 | 58.4 | 135  | 2 | H69202 | hypothetical protein  | 1.77e+01 |
| 23 | 52 | 58.4 | 489  | 2 | A47259 | corticosteroid bindin | 1.77e+01 |
| 22 | 52 | 58.4 | 3020 | 2 | A43932 | mycristatin           | 1.77e+01 |
| 21 | 52 | 58.4 | 177  | 2 | U5748  | coronatic acid synt   | 1.77e+01 |
| 20 | 51 | 57.3 | 521  | 2 | A4352  | conserved hypocheta   | 2.61e+00 |
| 19 | 51 | 57.3 | 774  | 2 | J6095  | hepatocyte nuclear fa | 2.61e+00 |
| 18 | 51 | 57.3 | 1747 | 2 | A45974 | collagen alpha 1(XIV) | 2.61e+00 |
| 17 | 51 | 57.3 | 89   | 2 | J01424 | hypothetical 9.8k pro | 3.81e+00 |
| 16 | 50 | 56.2 | 91   | 2 | S37486 | gene ms3 protein - m  | 3.81e+00 |
| 15 | 50 | 56.2 | 169  | 1 | RMSG2  | T-cell receptor gamma | 3.81e+00 |
| 14 | 50 | 56.2 | 179  | 2 | SS4801 | hypothetical protein  | 3.81e+00 |
| 13 | 50 | 56.2 | 217  | 2 | J09754 | hypothetical protein  | 3.81e+00 |
| 12 | 50 | 56.2 | 341  | 2 | JH0606 | thromboxane A2 recept | 3.81e+00 |
| 11 | 50 | 56.2 | 360  | 2 | S35453 | hypothetical protein  | 3.81e+00 |
| 10 | 50 | 56.2 | 386  | 2 | A3563  | NADH dehydrogenase (E | 3.81e+00 |
| 9  | 50 | 56.2 | 393  | 3 | A48357 | nonstructural protein | 3.81e+00 |
| 8  | 50 | 56.2 | 491  | 2 | UC2497 | cyclin E - mouse      | 3.81e+00 |
| 7  | 50 | 56.2 | 600  | 2 | A45452 | probable membrane pro | 3.81e+00 |
| 6  | 50 | 56.2 | 618  | 2 | SE9067 | probable glycine--trn | 3.81e+00 |
| 5  | 50 | 56.2 | 818  | 2 | PF9676 | hypothetical protein  | 3.81e+00 |
| 4  | 50 | 56.2 | 885  | 2 | S20621 | phycobilisome anchor  | 3.81e+00 |

## ALIGNMENTS

| ENTRY  | TITLE  | ORGANISM                                   | DATE                                       | RESULT       | 1           |
|--------|--|--|--|--------------|-------------|
| JC2358 | #type complete tumor-associated antigen , MAGE-1 - human | #formal_name Homo sapiens #common_name man | 20-Feb-1995 #sequence_revision 20-Feb-1995 | #text_change | 1E-04--1006 |

ACCESSIONS  
REFERENCE

| #authors   | #journal                     | #title      | #abstract |
|--|------------------------------|-------------|-----------|
| Ding, M.; Beck, R.J.; Keller, C.J.; Fenton, R.G. | Biophys. Res. Commun. (1994) | 202:549-555 |           |
| Cloning and analysis of MAGE-1-related genes.    |                              |             | 103350    |

```
##mol
##res
```

```
#experimental_source melanoma cell line DM150
GENETICS
#gene MAGE
removes
```

161-169  
SUMMARY

```
#region HLA-A1 binding #status predicted
#length 280 #molecular-weight 30932 #checksum 467
```

Query Match

|         |                  |               |           |         |   |
|---------|------------------|---------------|-----------|---------|---|
| Matches | 10; Conservative | 0; Mismatches | 1; Indels | 0; Gaps | 0 |
|---------|------------------|---------------|-----------|---------|---|

Db 92 C

QY 1 CILSCRAVI 11

| RESULT | 2 |
|--------|---|
| ENTRY  |   |

| ENRICHMENT | TITLE   |
|------------|---|
| 559044     | #lype complete                                    |
|            | hypothetical protein SPAC31A2.05c - fission yeast |
|            | (Schizosaccharomyces pombe)                       |

ORGANISM  
DATE

```

ACCESSIONS
      24 Jan 1990 *sequence_revision 15 Apr 1990 *rev_cchanged
      31-Oct-1997
      55964A

```

## REFERENCE

| #submisson | submitted to the EMBL Data Library, July 1995 |
|------------|---|
| #accession | S59664  |

```
##sta
##mol
```

```
##residues      1-1583 ##label DEV
##cross-references EMBL:Z50113 NID:q914878 PTD:q914883
```

## GENERAL

## GENETICS

```

#introns
SUMMARY      33/1: 98/2: 543/3: 699/3: 1294/2: 1339/3: 1558/3
#length 1583 #molecular-weight 180202 #checksum 4709
Query Match 68.5%; Score 61; DB 2; Length 1583;
Best Local Similarity 85.7%; Pred. No. 4,56e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1306 CILDCSF 1312
|1111111
|1 CILSCF 7

RESULT 3
ENTRY      VCMVJA      #type complete
TITLE      env polypeptide precursor - sheep pulmonary adenomatosis
VIRUS
ALTERNATE_NAMES coat polypeptide
CONTAINS      coat protein gp36; coat protein gp52
ORGANISM      formal name sheep pulmonary adenomatosis virus
DATE          31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change
05-Sep-1997

ACCESSIONS  E42740
REFERENCE    A42740
#authors    York, D.F.; Vigne, R.; Verwoerd, D.W.; Querat, G.
#journal    J. Virol. (1992) 66:4930-4939
#title      Nucleotide sequence of the jaagsiekte retrovirus, an
            exogenous and endogenous type D and B retrovirus of sheep
            and goats.
#cross-references M01D:92333675
#accession      E42740
#molecule_type genomic RNA
#residues       1-615 ##label YOR
#cross-references GB:M80216; MID:9331338; PID:9331342

GENETICS
#gene          env;
CLASSIFICATION #superfamily type A retrovirus env polypeptide
KEYWORDS      coat protein; glycoprotein; polypeptide; transmembrane
            protein.
FEATURE
#position 1-738
#0-378      #domain signal sequence #status predicted #label SIG
#379-615    #product coat protein gp52 #status predicted #label CP1
#379-402    #product coat protein gp36 #status predicted #label CP2
#403-615    #domain transmembrane #status predicted #label TM1
#555-571    #domain intracellular #status predicted #label INT
#108-127,178,219, #domain transmembrane #status predicted #label TM2
#binding-site carbohydrate (Asn) (covalent) #status
            predicted
SUMMARY      #length 615 #molecular-weight 69343 #checksum 8020
Query Match 65.2%; Score 58; DB 1; Length 615;
Best Local Similarity 60.0%; Pred. No. 1,60e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 323 CILMNCIRGV 332
|1111111
|1 CILSCFRAV 10

RESULT 4
ENTRY      S72705      #type complete
TITLE      mycocerosic acid synthase masA - Mycobacterium leprae
ALTERNATE_NAMES LepD170_C2_209 protein
ORGANISM      formal name Mycobacterium leprae
DATE          19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
10-Oct-1997

ACCESSIONS  S72705
REFERENCE    S72693
#authors    Smith, D.R.; Robison, K.
#submission submitted to the EMBL Data Library, November 1993
#description Mycobacterium leprae cosmid B1170.
#accession  S72705
#status     preliminary

```

```

#molecule_type DNA
#residues       1-2118 ##label SMI
#cross-references EMBL:000010; MID:9466780; PID:9466793

GENETICS
#start_codon  TTG
CLASSIFICATION #superfamily [acyl-carrier-protein] S-malonyltransferase
            homology; acyl carrier protein homology; short-chain
            alcohol dehydrogenase homology
FEATURE
536-816      #domain [acyl-carrier-protein] S-malonyltransferase
            homology #label AMT
1770-1954    #domain short-chain alcohol dehydrogenase homology
            #label SADH
SUMMARY      2038-2110      #domain acyl carrier protein homology #label ACPI
            #length 2118 #molecular-weight 226495 #checksum 3824
Query Match 65.2%; Score 58; DB 2; Length 2118;
Best Local Similarity 50.0%; Pred. No. 1,60e+00;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 1099 LIDACFOSVI 1108
|1111111
|2 ILSCFRAVI 11

RESULT 5
ENTRY      RRMEV      #type complete
TITLE      RNA-directed RNA polymerase (EC 2.7.7.48) - equine arteritis
            virus
ORGANISM      formal name equine arteritis virus
            host Equus caballus (domestic horse)
DATE          30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
13-Mar-1997

ACCESSIONS  A39925
REFERENCE    A39925
#authors    Den Boon, J.A.; Snijder, E.J.; Chirnside, E.D.; De Vries,
            A.A.F.; Horzinek, M.C.; Spaan, W.J.M.
#journal    J. Virol. (1991) 65:2910-2920
#title      Equine arteritis virus is not a togavirus but belongs to the
            coronaviruslike superfamily.
#cross-references M01D:91237805
#accession  A39925
#molecule_type genomic RNA
#residues  1-3175 ##label DEN
#cross-references EMBL:X53459
#note      a -1 ribosomal frameshift occurs between the codons AAC
            for 1727-Asn and CUG for 1728-Leu
REFERENCE    S10158
#authors    de Vries, A.A.F.; Chirnside, E.D.; Bredendiek, P.J.;
            Gravesstein, L.A.; Horzinek, M.C.; Spaan, W.J.M.
#journal    Nucleic Acids Res. (1990) 18:3241-3247
#title      All subgenomic mRNAs of equine arteritis virus contain a
            common leader sequence.
#cross-references M01D:90287699
#accession  S10158
#status     translation not shown
#molecule_type genomic RNA
#residues  1-17 ##label VRI
#cross-references EMBL:X52277

CLASSIFICATION #superfamily equine arteritis virus RNA-directed RNA
            polymerase
KEYWORDS      nucleotidyltransferase
SUMMARY      #length 3175 #molecular-weight 345277 #checksum 9571
Query Match 64.0%; Score 57; DB 1; Length 3175;
Best Local Similarity 60.0%; Pred. No. 2,42e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 344 CIDSCFROI 353
|1111111
|1 CILSCFRAV 10

```

```

RESULT 6
ENTRY OEOC03 #type complete
TITLE hypothetical 49.6 kd protein in asna 3' region - Escherichia
ORGANISM #formal_name Escherichia coli
DATE 17-May-1985 #sequence_revision 30-Sep-1997 #text_change
ACCESSION B65178; A04443
REFERENCE B65178; A04443
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Butland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Roe, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references M01D:97426517
#accession B65178
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-427 #label BLAT
#cross-references GB:AE000451; GB:U00096; NID:92367272; PID:92367274;
UMCP:B3745
REFERENCE A91504
#authors Buhk, H.J.; Messer, W.
#journal Gene (1983) 24:265-279
#title The replication origin region of Escherichia coli: nucleotide
sequence and functional units.
#cross-references M01D:84059088
#accession A04443
#molecule_type DNA
#residues 128-427 #label Buh
GENETICS
#map_position 84 min
SUMMARY #length 427 #molecular-weight 49625 #checksum 4786
Query Match 62.9%; Score 56; DB 1; Length 427;
Best Local Similarity 66.7%; Pred. No. 3.63e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 339 LASCRAV 347
QY 3 LASCRAV 11
RESULT 7
ENTRY A32893 #type complete
TITLE foliitropin beta chain precursor - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 29-Jan-1990 #sequence_revision 29-Jan-1990 #text_change
ACCESSION A32893; A40060
REFERENCE A32893
#authors Charib, S.D.; Roy, A.; Wiernan, M.E.; Chin, W.W.
#journal DNA (1989) 8:339-349
#title Isolation and characterization of the gene encoding the
beta-subunit of rat follicle-stimulating hormone.
#cross-references M01D:89356263
#accession A32893
#status preliminary
#molecule_type DNA
#residues 1-130 #label GHA
#cross-references GB:M27044; GB:M27048; NID:9204179; PID:9204181
REFERENCE A40060
#authors Maurer, R.A.
#journal Mol. Endocrinol. (1987) 1:717-723
#title Molecular cloning and nucleotide sequence analysis of
rat follicle stimulating hormone.
#cross-references M01D:91042555
#accession A40060
#status preliminary

```

```

#molecule_type mRNA
#residues 1-130 #label MAU
#cross-references GB:M36804
CLASSIFICATION #superfamily pituitary glycoprotein hormone beta chain
FEATURE
22-47,36-70,39-101,
51-123,85-113,
103-106
SUMMARY #length 130 #molecular-weight 14814 #checksum 2096
Query Match 60.7%; Score 54; DB 2; Length 130;
Best Local Similarity 70.0%; Pred. No. 8.10e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 8 CILMCLRAV 17
QY 1 CILMCLRAV 10
RESULT 8
ENTRY C69102 #type complete
TITLE DNA mismatch recognition protein Muls - Methanobacterium
thermautotrophicum (strain Delta H)
ORGANISM #formal_name Methanobacterium thermautotrophicum
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSION C69102
REFERENCE A69000
#authors Smith, D.R.; Doucette-Stamm, L.A.; DeLoughery, C.; Lee, H.;
Dubols, J.; Aldredge, T.; Bashirzadeh, R.; Blakey, D.;
Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
Lumm, W.; Pouchet, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Carnio,
A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
McDougal, S.; Shiner, G.; Goyal, A.; Pietrovski, S.;
Church, G.M.; Daniels, C.J.; Mac, J.; Rice, P.; Noelling,
J.; Reeve, J.N.
#journal J. Bacteriol. (1997) 179:7135-7155
#title Complete genome sequence of Methanobacterium
thermautotrophicum Delta H: functional analysis and
comparative genomics.
#cross-references M01D:98037514
#accession C69102
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-647 #label MTH
#cross-references GB:AE000666
#experimental_source strain Delta H
GENETICS
#gene MTH1762
#start_codon TTG
SUMMARY #length 647 #molecular-weight 73592 #checksum 610
Query Match 60.7%; Score 54; DB 2; Length 647;
Best Local Similarity 60.0%; Pred. No. 8.10e+00;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 487 CALMCLRAV 496
QY 1 CILMCLRAV 10
RESULT 9
ENTRY I55975 #type fragment
TITLE X/Y protein - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
ACCESSION I55975
REFERENCE I55975
#authors Aegerter-Shaw, M.; Cole, J.L.; Klickstein, L.B.; Wong, W.W.;
Fearon, D.T.; Lally, P.A.; Weis, J.H.

```

#journal J. Immunol. (1987) 138:3488-3494  
 #title Expansion of the complement receptor gene family: identification in the mouse of two new genes related to the CR1 and CR2 gene family.  
 #cross-references M01D:87196375  
 #accession I55975  
 #status Preliminary; translated from GB/EMBL/DBJ  
 #molecule-type mRNA  
 #residues 1-330 ##label RES  
 #cross-references GB:M61179; NID:9202427; PID:9202428  
 CLASSIFICATION #superfamily complement factor H repeat homology  
 FEATURE #domain complement factor H repeat homology #label FH4  
 36-92 #domain complement factor H repeat homology #label FH1  
 188-244 #domain complement factor H repeat homology #label FH2  
 249-306 #domain complement factor H repeat homology #label FH02  
 SUMMARY #length 330 #checksum 9931

Query Match 59.6%; Score 53; DB 2; Length 330;  
 Best Local Similarity 66.7%; Pred. No. 1.20e+01;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 89 LPSCFRAVI 97  
 OY 3 LESCRAVI 11

RESULT 10  
 ENTRY JCA938 #type complete  
 TITLE hepatocyte nuclear factor 4C - human  
 ORGANISM #formal\_name Homo sapiens #common\_name man  
 DATE 22-Oct-1996 #sequence\_revision 01-Nov-1996 #text\_change 31-Oct-1997  
 ACCESSIONS JCA938  
 REFERENCE JCA936  
 #authors Kritis, A.A.; Agyrokastritis, A.; Moschonas, N.K.; Power, S.; Katrakili, N.; Zannis, V.I.; Cereghini, S.; Tallanidis, I.

#journal Gene (1996) 173:275-280  
 #title Isolation and characterization of a third isoform of human hepatocyte nuclear factor 4.  
 #accession JCA938  
 #status Preliminary  
 #molecule-type mRNA  
 #residues 1-408 ##label KRI  
 #cross-references EMBL:X87872; NID:91595753; PID:91595754  
 #experimental\_source liver  
 ##note nucleic acid sequence not complete  
 COMMENT This protein is one of the positive regulators of liver-specific genes.

GENETICS  
 #gene hHNF-4C  
 #map\_position 20  
 CLASSIFICATION #superfamily unassigned erba-related proteins; erba transforming protein homology  
 KEYWORDS zinc finger  
 FEATURE 49-288  
 SUMMARY #domain erba transforming protein homology #label ERBA  
 #length 408 #molecular-weight 45578 #checksum 5015

Query Match 59.6%; Score 53; DB 2; Length 408;  
 Best Local Similarity 66.7%; Pred. No. 1.20e+01;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 106 CRUKCFRA 114  
 OY 1 CILSCFRA 9

RESULT 11  
 ENTRY A30550 #type complete  
 TITLE complement C3b/C4b receptor precursor - mouse  
 ORGANISM #formal\_name Mus musculus #common\_name house mouse  
 DATE 03-Mar-1989 #sequence\_revision 03-Mar-1989 #text\_change

ACCESSIONS 12-May-1995  
 REFERENCE A30550  
 #authors Paul, M.S.; Aegeerter, M.; O'Brien, S.E.; Kurtz, C.B.; Wels, J.H.

#journal J. Immunol. (1989) 142:582-589  
 #title The murine complement receptor gene family. Analysis of mcrv gene products and their homology to human CR1.  
 #cross-references M01D:8903944  
 #accession A30550  
 #status Preliminary  
 #molecule-type mRNA  
 #residues 1-433 ##label PAU  
 CLASSIFICATION #superfamily complement factor H repeat homology  
 FEATURE 42-98  
 103-160 #domain complement factor H repeat homology #label FH1  
 165-231 #domain complement factor H repeat homology #label FH2  
 237-293 #domain complement factor H repeat homology #label FH3  
 299-355 #domain complement factor H repeat homology #label FH5  
 SUMMARY #length 433 #molecular-weight 48344 #checksum 2181

Query Match 59.6%; Score 53; DB 2; Length 433;  
 Best Local Similarity 66.7%; Pred. No. 1.20e+01;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 290 LPSCFRAVI 298  
 OY 3 LESCRAVI 11

RESULT 12  
 ENTRY S55631 #type complete  
 TITLE virion protein kinase 36 - equine herpesvirus 2  
 ORGANISM #formal\_name equine herpesvirus 2  
 DATE 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 09-Sep-1997  
 ACCESSIONS S55631  
 REFERENCE S55594  
 #authors Telford, E.A.R.; Watson, M.S.; Alrd, H.C.; Perry, J.; Davidson, A.J.

#journal J. Mol. Biol. (1995) 249:520-528  
 #title The DNA sequence of equine herpesvirus 2.  
 #accession S55631  
 #status Preliminary; nucleic acid sequence not shown; translation not shown  
 ##molecule-type DNA  
 #residues 1-438 ##label TEL  
 #cross-references GB:U20824; NID:9695172; PID:9695209  
 ##note the nucleotide sequence was submitted to the EMBL Data Library, February 1995

SUMMARY #length 438 #molecular-weight 49385 #checksum 5511

Query Match 59.6%; Score 53; DB 2; Length 438;  
 Best Local Similarity 62.5%; Pred. No. 1.20e+01;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 265 CVALRCFR 272  
 OY 1 CILSCFR 8

RESULT 13  
 ENTRY A43519 #type complete  
 TITLE complement receptor CR1 precursor - mouse  
 ORGANISM #formal\_name Mus musculus #common\_name house mouse  
 DATE 28-Oct-1992 #sequence\_revision 30-Jan-1993 #text\_change 12-May-1995  
 ACCESSIONS A43519  
 REFERENCE A43519  
 #authors Paul, M.S.; Aegeerter, M.; Cepek, K.; Miller, M.D.; Wels, J.H.  
 #journal J. Immunol. (1990) 144:1988-1996  
 #title The murine complement receptor gene family. The genomic and

## Transcriptional complexity of the Cry and Cry-ps genes.

#cross-references WUID:90211600  
#accession A43519  
#status Preliminary  
#molecule\_type DNA  
#residues 1-440 #label PAU  
#cross-references GB:M4164  
#note the authors translated the codon GGC for residue 21 as Ala, and CAG for residue 121 as Glu

CLASSIFICATION #superfamily complement factor H repeat homology

FEATURE  
42-98 #domain complement factor H repeat homology #label FH1  
103-160 #domain complement factor H repeat homology #label FH2  
165-231 #domain complement factor H repeat homology #label FH3  
237-293 #domain complement factor H repeat homology #label FH4  
299-355 #domain complement factor H repeat homology #label FH5  
SUMMARY #length 440 #molecular-weight 49074 #checksum 3752

Query Match 59.6%; Score 53; DB 2; Length 440;  
Best Local Similarity 66.7%; Pred. No. 1.20e+01;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 290 LPSCFKGVI 298  
QY 3 LESCFAVI 11

RESULT 14  
ENTRY JC4936 #type complete  
TITLE hepatocyte nuclear factor 4A - human  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 22-Oct-1996 #sequence\_revision 01-Nov-1996 #text\_change 31-Oct-1997

ACCESSIONS JC4936  
REFERENCE JC4936  
#authors Krutts, A.A.; Argyrokastritis, A.; Moschonas, N.K.; Power, S.; Katriakili, N.; Zannis, V.I.; Cereghini, S.; Talianidis, I.

#journal Gene (1996) 173:275-280  
#title Isolation and characterization of a third isoform of human hepatocyte nuclear factor 4.  
#accession JC4936  
#status Preliminary  
#molecule\_type mRNA  
#residues 1-455 #label KRI  
#cross-references EMBL:X87870; NID:G1595751; PID:G1595752  
#experimental\_source liver  
#note nucleic acid sequence is not complete in this paper  
COMMENT This protein is one of the positive regulators of liver-specific genes.

GENETICS  
#gene hNF-4A  
#map\_position 20  
CLASSIFICATION #superfamily unassigned erba-related proteins; erba  
transforming protein homology  
KEYWORDS zinc finger  
FEATURE 49-288  
SUMMARY #domain erba transforming protein homology #label ERBA  
#length 455 #molecular-weight 50565 #checksum 3625

Query Match 59.6%; Score 53; DB 2; Length 455;  
Best Local Similarity 66.7%; Pred. No. 1.20e+01;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 106 CRLKCFRA 114  
QY 1 CILSCFRA 9

RESULT 15  
ENTRY A36471 #type complete  
TITLE transcription factor HNF-4 - rat  
ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat

DATE 19-Apr-1991 #sequence\_revision 19-Apr-1991 #text\_change 12-Sep-1997

ACCESSIONS A36471  
REFERENCE A36471  
#authors Sladek, F.M.; Zhong, W.; Lai, E.; Darnell Jr., J.E.  
#journal Genes Dev. (1990) 4:2353-2365  
#title Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily.  
#cross-references WUID:91122637  
#accession A36471  
#status Preliminary  
#molecule\_type mRNA  
#residues 1-455 #label SLA  
#cross-references GB:X57133; NID:956371; PID:956372  
CLASSIFICATION #superfamily unassigned erba-related proteins; erba  
transforming protein homology  
KEYWORDS DNA binding; transcription regulation; zinc finger  
FEATURE 49-288  
SUMMARY #domain erba transforming protein homology #label ERBA  
#length 455 #molecular-weight 50530 #checksum 1464

Query Match 59.6%; Score 53; DB 2; Length 455;  
Best Local Similarity 66.7%; Pred. No. 1.20e+01;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 106 CRLKCFRA 114  
QY 1 CILSCFRA 9

Search completed: Fri Sep 11 12:46:37 1998  
Job time : 34 secs.

**THIS PAGE BLANK (USPTO)**





DR EMBL: M77481; G416115; -  
 DR EMBL: U82672; G2078527; -  
 DR MIM: 300016; -  
 KM ANTIGEN: MULTIGENE FAMILY: POLYMORPHISM: TUMOR ANTIGEN.  
 FT VARIANT 32 32  
 FT DOMAIN 33 36  
 FT MUTAGEN 163 163  
 FT MUTAGEN 169 169  
 FT CONFLICT 72 72  
 SO SEQUENCE 309 AA; 34342 MM; E6CB1300 CRC32;.

Query Match  
 Best Local Similarity 75.38; Score 67; DB 1; Length 309;  
 Matches 10; Conservative 0; Pred. No. 3.71e-03;  
 Mismatches 1; Indels 0; Gaps 0;

Db 92 CILSCFRAY 102  
 1 CILSCFRAY 11

RESULT 2  
 ID VA45.SCHPO STANDARD; PRT: 1583 AA.  
 AC 009725;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DE HYPOTHETICAL 180.2 KD PROTEIN C1A2.05C IN CHROMOSOME 1.  
 GN SPAC31A2.05C  
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972;  
 RA DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAN M.A., WALSH S.V.;  
 RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 DR EMBL: Z50113; G914883; -  
 KW HYPOTHETICAL PROTEIN: TRANSMEMBRANE.  
 FT TRANSMEM 319 339  
 FT TRANSMEM 633 653  
 FT TRANSMEM 764 784  
 FT TRANSMEM 1014 1034  
 FT TRANSMEM 1216 1236  
 FT TRANSMEM 1452 1472  
 SO SEQUENCE 1583 AA; 180203 MM; 9617C75D CRC32;

Query Match  
 Best Local Similarity 68.58; Score 61; DB 1; Length 1583;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1306 CILSCF 1312  
 1 CILSCF 7

RESULT 3  
 ID ENV.JSRV STANDARD; PRT: 615 AA.  
 AC P31621;  
 DT 01-JUL-1993 (REL. 26, CREATED)  
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
 DE ENV POLYPROTEIN PRECURSOR (COAT POLYPROTEIN) (CONTAINS: COAT PROTEIN  
 GN GP52; COAT PROTEIN GP36).  
 OS SHEEP PULMONARY ADENOMATOSIS VIRUS (JAGSIEKTE SHEEP RETROVIRUS)  
 OS (JUVAV).  
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 OC LENTIVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92333675.  
 RA YÖRK D.E., VIGNE R., VERMOERD D.W., QUERAT G.;  
 RL J. VIROL. 66:4930-4939(1992).

DR EMBL: M80216; G331342; -  
 DR PIR: A42740; VCMVIA.  
 KM COAT PROTEIN; GLYCOPROTEIN; POLYPROTEIN; TRANSMEMBRANE.  
 FT PROPEP 1 79  
 FT CHAIN 80 378  
 FT CHAIN 379 615  
 FT DOMAIN 80 378  
 FT TRANSMEM 379 402  
 FT DOMAIN 403 615  
 FT CARBOHYD 108 108  
 FT CARBOHYD 127 127  
 FT CARBOHYD 178 178  
 FT CARBOHYD 219 219  
 FT CARBOHYD 275 275  
 FT CARBOHYD 319 319  
 FT CARBOHYD 420 420  
 FT CARBOHYD 479 479  
 FT CARBOHYD 500 500  
 SO SEQUENCE 615 AA; 69343 MM; 78B74F63 CRC32;

Query Match  
 Best Local Similarity 65.28; Score 58; DB 1; Length 615;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 323 CILNCIRGV 332  
 1 CILNCIRGV 10

RESULT 4  
 ID RPOL.EAV STANDARD; PRT: 3175 AA.  
 AC P19811;  
 DT 01-FEB-1991 (REL. 17, CREATED)  
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DE POL POLYPROTEIN (CONTAINS: RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48);  
 DE HELICASE; PROTEASE (EC 3.4.21.-)) (ORF1A/1B).  
 OS EQUINE ARTERITIS VIRUS (EAV).  
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; TOGAVIRIDAE;  
 OC ARTERIVIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BUCYRUS;  
 RX MEDLINE: 91237805.  
 RA DEN BOON J.A., SNIDER E.J., CHIRNSIDE E.D., DE VRIES A.A.F.,  
 RA HORZINKER M.C., SPAN W.J.M.;  
 RL J. VIROL. 65:2910-2920(1991).  
 [2]  
 RN RN  
 RP SEQUENCE OF 1-17 FROM N.A.  
 RC STRAIN-BUCYRUS;  
 RX MEDLINE: 90287699.  
 RA DE VRIES A.A.F., CHIRNSIDE E.D., BREDEMEER P.J., GRAVESTEN L.A.,  
 RA HORZINKER M.C., SPAN W.J.M.;  
 RL NUCLEIC ACIDS RES. 18:3241-3247(1990).  
 CC -1- FUNCTION: RNA-DIRECTED RNA POLYMERASE & POSSIBLE HELICASE. A  
 CC SUGGESTED. ALSO CONTAINS A PROTEASE DOMAIN.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -  
 CC N PYROPHOSPHATE + RNA(N).  
 CC -1- A RIBOSOMAL FRAMESHIFT OCCURS BETWEEN THE CODONS FOR 1727-ASN AND  
 CC 1728-LEU.  
 CC -1- SIMILARITY: WITH THE POLYMERASE PROTEIN OF OTHER CORONAVIRUSES AND  
 CC OF TOROVIRUSES.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S32.  
 DR EMBL: X53459; G62066; ALT\_FRAME.  
 DR EMBL: X52277; G59071; -  
 DR PIR: S10158; S10158.  
 DR PIR: A39925; RRTREV.  
 KM RNA-DIRECTED RNA POLYMERASE; HELICASE; ATP-BINDING; HYDROLASE;  
 KM SERINE PROTEASE.  
 FT CHAIN 1 1727  
 FT CHAIN 1728 3175  
 ORF1A.  
 ORF1B.

FT DOMAIN 1080 1220 TRYPSIN-LIKE SERINE PROTEASE.  
 FT DOMAIN 1218 1506 HELICASE.  
 FT DOMAIN 2098 2306 POLYMERASE.  
 FT ACT\_SITE 1103 1103 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1129 1129 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1184 1184 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ZN\_FING 2368 2414 BY SIMILARITY.  
 FT NP\_BIND 2528 2535 ATP (BY SIMILARITY).  
 SO SEQUENCE 3175 AA; 345275 MW; E87D3E59 CRC32;

Query Match 64.08; Score 57; DB 1; Length 3175;  
 Best Local Similarity 60.08; Pred. No. 5.44e-01;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 344 CLDSCFRGI 353  
 QY 1 CLDSCFRAY 10

RESULT 5  
 ID YIED.ECOLI STANDARD; PRT; 298 AA.  
 AC P03818;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 34.8 KD PROTEIN IN ASNA-KUP INTERGENIC REGION.  
 GN YIED.  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RX MEDLINE: 84059088.  
 RA BUHK H.-J., MESSER W.;  
 RL GENE 24:265-279(1983).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE: 93315143.  
 RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;  
 RL GENOMICS 16:551-561(1993).  
 RN [3]  
 RP SEQUENCE OF 253-298 FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE: 82059491.  
 RA NAKAMURA M., YAMADA M., HIROTA Y., SUGIMOTO K., OKA A., TAKANAMI M.;  
 RL NUCLEIC ACIDS RES. 9:4669-4676(1981).  
 DR EMBL: K00826; NOT\_ANNOTATED\_CDS.  
 DR EMBL: V00263; NOT\_ANNOTATED\_CDS.  
 DR EMBL: L10328; G290594.  
 DR EMBL: AE000451; G1790184; .  
 DR PIR: A04443; Q0EC03.  
 DR ECOGENE: EG11365; YIED.  
 KW HYPOTHETICAL PROTEIN.  
 SO SEQUENCE 298 AA; 34754 MW; ADAF6165 CRC32;

Query Match 62.98; Score 56; DB 1; Length 298;  
 Best Local Similarity 66.78; Pred. No. 8.73e-01;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 210 LASCFRAM 218  
 QY 3 LASCFRAY 11

RESULT 6  
 ID FSHB.RAT STANDARD; PRT; 130 AA.  
 AC P18427;  
 DT 01-NOV-1990 (REL. 16, CREATED)  
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE FOLLITROPIN BETA CHAIN PRECURSOR (FOLLICLE-STIMULATING HORMONE)  
 DE (FSH-B).

GN FSHB.  
 OS RATUS NORVEGICUS (RAT).  
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HOLTZMAN;  
 RX MEDLINE: 91042555.  
 RA MAURER R.A.;  
 RL MOL. ENDOCRINOL. 1:717-723(1987).

Query Match 60.78; Score 54; DB 1; Length 130;  
 Best Local Similarity 70.08; Pred. No. 2.21e+00;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 8 CILMLCLAY 17  
 QY 1 CILMCLAY 10

RESULT 7  
 ID HNF4.RAT STANDARD; PRT; 465 AA.  
 AC P22449;  
 DT 01-AUG-1991 (REL. 19, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HEPATOCYTE NUCLEAR FACTOR 4 (TRANSCRIPTION FACTOR HNF-4)  
 DE (TRANSCRIPTION FACTOR 14).  
 GN TCF4 OR HNF4 OR HNF-4.  
 OS RATUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 91122637.

RA SLADER F.M., ZHONG W., LAI E., DARNELL J.E. JR.;  
 RL GENES DEV. 4:2353-2365(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 RC STRAIN-MISTAR; TISSUE-LIVER;  
 RA MEDLINE: 92305063.  
 RA HATA S., TSUKAMOTO T., OSUMI T.;  
 RL BIOCHIM. BIOPHYS. ACTA 1131:211-213(1992).  
 CC -1- FUNCTION: TRANSCRIPTIONALLY CONTROLLED TRANSCRIPTION FACTOR. BINDS  
 TO DNA SITES REQUIRED FOR THE TRANSCRIPTION OF ALPHA 1-  
 ANTIPIPSIN, APOLOPROTEIN CIII AND TRANSFERRIN GENES. MAY BE  
 ESSENTIAL FOR DEVELOPMENT OF THE LIVER, KIDNEY AND INTESTINE.  
 CC -1- SUBUNIT: HOMODIMERIZATION IS REQUIRED FOR HNF-4 TO BIND TO ITS  
 RECOGNITION SITE.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: LIVER, KIDNEY AND INTESTINE.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE  
 SPLICING.  
 CC -1- SIMILARITY: TO OTHER MEMBERS OF THE STEROID/THYROID/RETINOIC  
 NUCLEAR HORMONE RECEPTORS.  
 DR EMBL: D10554; G220773; -  
 DR EMBL: X57133; G56372; -  
 DR PIR: A36471; A36471.  
 DR HSSP: P10826; IHRA.  
 DR TRANSFAC: T00372; -  
 DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.  
 KW RECEPTOR: TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN;  
 KM ZINC-FINGER; LIVER; ALTERNATIVE SPLICING.  
 FT MOD\_RRS 71 71  
 FT DN\_BIND 51 116 C4-TYPE ZINC FINGERS (TWO).  
 FT ZN\_FING 87 71 C4-TYPE.  
 FT ZN\_FING 87 111 C4-TYPE.  
 FT VARSPLIC 409 419 CEMPRGQA -> S (IN SHORT FORM).  
 FT CONFLICT 171 171 K -> R (IN REF. 1).  
 FT CONFLICT 174 174 N -> S (IN REF. 1).  
 SQ SEQUENCE 465 AA; 51695 MW; 15DDEF CRC32;  
 Query Match 59.6%; Score 53; DB 1; Length 465;  
 Best Local Similarity 66.7%; Pred. No. 3,48e+00;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Db 106 CRKCFRA 114  
 1 CILSCFRA 9  
 QY 1 CILSCFRA 9  
 RESULT 8  
 ID HNF4\_HUMAN STANDARD; PRT; 465 AA.  
 AC P41235;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HEPATOCYTE NUCLEAR FACTOR 4 (TRANSCRIPTION FACTOR HNF-4)  
 DE (TRANSCRIPTION FACTOR 14).  
 DE TCF14 OR HNF4.  
 GN HOMO SAPIENS (HUMAN).  
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 95011627.  
 RA CHARTIER F.L., BOSSU J.-P., LAUDET V., FRUCHART J.-C., LAINE B.;  
 RL GENE 147:269-272(1994).  
 CC -1- FUNCTION: TRANSCRIPTIONALLY CONTROLLED TRANSCRIPTION FACTOR. BINDS  
 TO DNA SITES REQUIRED FOR THE TRANSCRIPTION OF ALPHA 1-  
 ANTIPIPSIN, APOLOPROTEIN CIII AND TRANSFERRIN GENES. MAY BE  
 ESSENTIAL FOR DEVELOPMENT OF THE LIVER, KIDNEY AND INTESTINE.  
 CC -1- SUBUNIT: HOMODIMERIZATION IS REQUIRED FOR HNF-4 TO BIND TO ITS  
 RECOGNITION SITE.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE  
 SPLICING.

CC -1- SIMILARITY: TO OTHER MEMBERS OF THE STEROID/THYROID/RETINOIC  
 NUCLEAR HORMONE RECEPTORS.  
 DR EMBL: X76930; G575253; -  
 DR MIM: 600281; -  
 DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.  
 KW RECEPTOR: TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN;  
 KM ZINC-FINGER; LIVER; ALTERNATIVE SPLICING.  
 FT DN\_BIND 51 116 C4-TYPE ZINC FINGERS (TWO).  
 FT ZN\_FING 87 71 C4-TYPE.  
 FT ZN\_FING 87 111 C4-TYPE.  
 FT VARSPLIC 409 419 CEMPRGQA -> S (IN SHORT FORM).  
 SQ SEQUENCE 465 AA; 51754 MW; 1FD7E232 CRC32;  
 Query Match 59.6%; Score 53; DB 1; Length 465;  
 Best Local Similarity 66.7%; Pred. No. 3,48e+00;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Db 106 CRKCFRA 114  
 1 CILSCFRA 9  
 QY 1 CILSCFRA 9  
 RESULT 9  
 ID HNF4\_MOUSE STANDARD; PRT; 465 AA.  
 AC P49698;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HEPATOCYTE NUCLEAR FACTOR 4 (TRANSCRIPTION FACTOR HNF-4)  
 DE (TRANSCRIPTION FACTOR 14).  
 DE TCF14 OR HNF4 OR HNF-4.  
 GN MUS MUSCULUS (MOUSE).  
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B6/CBA; TISSUE-LIVER;  
 RX MEDLINE: 95092794.  
 RA HATA S., INOUE T., KOSUGA K., NAKASHIMA T., TSUKAMOTO T., OSUMI T.;  
 RL BIOCHIM. BIOPHYS. ACTA 1260:55-61(1995).  
 CC -1- FUNCTION: TRANSCRIPTIONALLY CONTROLLED TRANSCRIPTION FACTOR. BINDS  
 TO DNA SITES REQUIRED FOR THE TRANSCRIPTION OF ALPHA 1-  
 ANTIPIPSIN, APOLOPROTEIN CIII AND TRANSFERRIN GENES. MAY BE  
 ESSENTIAL FOR DEVELOPMENT OF THE LIVER, KIDNEY AND INTESTINE.  
 CC -1- SUBUNIT: HOMODIMERIZATION IS REQUIRED FOR HNF-4 TO BIND TO ITS  
 RECOGNITION SITE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE  
 SPLICING.  
 CC -1- SIMILARITY: TO OTHER MEMBERS OF THE STEROID/THYROID/RETINOIC  
 NUCLEAR HORMONE RECEPTORS.  
 DR EMBL: D29015; G511934; -  
 DR MGD: MGI:109128; TCF14.  
 DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.  
 KW RECEPTOR: TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN;  
 KM ZINC-FINGER; LIVER; ALTERNATIVE SPLICING.  
 FT DN\_BIND 51 116 C4-TYPE ZINC FINGERS (TWO).  
 FT ZN\_FING 87 71 C4-TYPE.  
 FT ZN\_FING 87 111 C4-TYPE.  
 FT VARSPLIC 409 419 CEMPRGQA -> S (IN SHORT FORM).  
 SQ SEQUENCE 465 AA; 51755 MW; 097865A9 CRC32;  
 Query Match 59.6%; Score 53; DB 1; Length 465;  
 Best Local Similarity 66.7%; Pred. No. 3,48e+00;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Db 106 CRKCFRA 114  
 1 CILSCFRA 9  
 QY 1 CILSCFRA 9  
 RESULT 10  
 ID NUC2\_SCHPO STANDARD; PRT; 665 AA.  
 AC P49698;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HEPATOCYTE NUCLEAR FACTOR 4 (TRANSCRIPTION FACTOR HNF-4)  
 DE (TRANSCRIPTION FACTOR 14).  
 DE TCF14 OR HNF4 OR HNF-4.  
 GN MUS MUSCULUS (MOUSE).  
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B6/CBA; TISSUE-LIVER;  
 RX MEDLINE: 95092794.  
 RA HATA S., INOUE T., KOSUGA K., NAKASHIMA T., TSUKAMOTO T., OSUMI T.;  
 RL BIOCHIM. BIOPHYS. ACTA 1260:55-61(1995).  
 CC -1- FUNCTION: TRANSCRIPTIONALLY CONTROLLED TRANSCRIPTION FACTOR. BINDS  
 TO DNA SITES REQUIRED FOR THE TRANSCRIPTION OF ALPHA 1-  
 ANTIPIPSIN, APOLOPROTEIN CIII AND TRANSFERRIN GENES. MAY BE  
 ESSENTIAL FOR DEVELOPMENT OF THE LIVER, KIDNEY AND INTESTINE.  
 CC -1- SUBUNIT: HOMODIMERIZATION IS REQUIRED FOR HNF-4 TO BIND TO ITS  
 RECOGNITION SITE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE  
 SPLICING.  
 CC -1- SIMILARITY: TO OTHER MEMBERS OF THE STEROID/THYROID/RETINOIC  
 NUCLEAR HORMONE RECEPTORS.  
 DR EMBL: D29015; G511934; -  
 DR MGD: MGI:109128; TCF14.  
 DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.  
 KW RECEPTOR: TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN;  
 KM ZINC-FINGER; LIVER; ALTERNATIVE SPLICING.  
 FT DN\_BIND 51 116 C4-TYPE ZINC FINGERS (TWO).  
 FT ZN\_FING 87 71 C4-TYPE.  
 FT ZN\_FING 87 111 C4-TYPE.  
 FT VARSPLIC 409 419 CEMPRGQA -> S (IN SHORT FORM).  
 SQ SEQUENCE 465 AA; 51755 MW; 097865A9 CRC32;  
 Query Match 59.6%; Score 53; DB 1; Length 465;  
 Best Local Similarity 66.7%; Pred. No. 3,48e+00;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Db 106 CRKCFRA 114  
 1 CILSCFRA 9  
 QY 1 CILSCFRA 9

AC P10505;  
 DT 01-JUL-1989 (REL. 11, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE NUCLEAR SCAFFOLD-LIKE PROTEIN P76.  
 GN NUC2 OR SPAC17C9.01C.  
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
 OC EUKARYOTA: FUNGI; ASCOMYCOTINA; HEMIASCOCYCTES.  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TS MUTANT NUC2-663;  
 RX MEDLINE; 88198361.  
 RA HIRANO M., HIRAKA Y., YANAGIDA M.;  
 RL J. CELL BIOL. 106:1171-1183(1988).  
 RM [2]  
 RP REVISION TO 649.  
 RA YANAGIDA M.;  
 RL SUBMITTED (MAR-1989) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE OF 1-557 FROM N.A.  
 RC STRAIN-972;  
 RA MURPHY L., MCDUGALL R., JONES L., SIMPSON I., MCNEIL A., HARRIS D.,  
 RA BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;  
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [4]  
 RP DOMAINS.  
 RX MEDLINE; 90124640.  
 RA HIRANO M., KINOSHITA N., MORIKAWA K., YANAGIDA M.;  
 RL CELL 60:319-328(1990).  
 CC -1- FUNCTION: NUC2 INTERACTS WITH SPINDLE APPARATUS, CHROMOSOMES,  
 OR NUCLEAR ENVELOPE, AND INTERCONNECT NUCLEAR AND CYTOSKELETAL  
 CC FUNCTIONS IN MITOSIS, SO THE ELONGATION OF THE SPINDLE IN ANAPHASE  
 CC IS BLOCKED.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC CC  
 CC EMBL; X07693; G4997;  
 DR EMBL; 273099; E241755;  
 DR PIR; A30185; A30185;  
 KM CELL DIVISION; CELL CYCLE; MITOSIS; REPEAT; TPR DOMAIN;  
 KM NUCLEAR PROTEIN.  
 FT REPEAT 118 151 TPR 1.  
 FT DNA BIND 191 257  
 FT REPEAT 332 365  
 FT REPEAT 366 399 TPR 2.  
 FT REPEAT 400 433 TPR 3.  
 FT REPEAT 434 467 TPR 4.  
 FT REPEAT 468 501 TPR 5.  
 FT REPEAT 502 535 TPR 6.  
 FT REPEAT 536 569 TPR 7.  
 FT REPEAT 570 603 TPR 8.  
 FT REPEAT 604 637 TPR 9.  
 FT REPEAT 604 637 TPR 10.  
 FT MUTAGEN 504 504  
 FT CONFLICT 440 440  
 SQ SEQUENCE 665 AA; 76171 MW; D46B3C8 CRC32;  
 Query Match 59.6%; Score 53; DB 1; Length 665;  
 Best Local Similarity 71.4%; Pred. No. 3,48e+00;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 DB 435 CILSCF 441  
 OY 1 CILSCF 7  
 RESULT 11  
 ID GLI4\_XENLA STANDARD; PRT: 1361 AA.  
 AC 091661;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ZINC FINGER PROTEIN GLI4 (NEURAL SPECIFIC DNA BINDING PROTEIN XGLI4).  
 GN GLI4.  
 OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).

OC EUKARYOTA: METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MARINE J.C., BELLEFROID E.J., PENDEVILLE H., MARTIAL J.A.,  
 RA PIELER T.;  
 RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: HAS AN ESSENTIAL ROLE IN THE EARLY EMBRYONIC PATTERNING  
 CC OF MESODERM AND NEUROECTODERM.  
 CC CC  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- SIMILARITY: TO THE GLI-RELATED GROUP OF C2H2-TYPE ZINC-FINGERS  
 CC PROTEINS.  
 DR EMBL; U42462; G1150838;  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_4.  
 KM ZINC-FINGER; METAL-BINDING; DNA-BINDING; TRANSCRIPTION REGULATION;  
 KM NUCLEAR PROTEIN.  
 FT DOMAIN 289 441 ZINC-FINGERS.  
 FT ZN\_FING 289 314 C2H2-TYPE.  
 FT ZN\_FING 322 349 C2H2-TYPE.  
 FT ZN\_FING 355 379 C2H2-TYPE.  
 FT ZN\_FING 385 410 C2H2-TYPE.  
 FT ZN\_FING 416 441 C2H2-TYPE.  
 SQ SEQUENCE 1361 AA; 149554 MW; 70E6495C CRC32;  
 Query Match 59.6%; Score 53; DB 1; Length 1361;  
 Best Local Similarity 55.6%; Pred. No. 3,48e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 DB 357 CTFEGCFRA 365  
 OY 1 CILSCFRA 9  
 RESULT 12  
 ID CID\_DROME STANDARD; PRT: 1377 AA.  
 AC P19538;  
 DT 01-NOV-1990 (REL. 16, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE CUBITUS INTERRUPTUS DOMINANT PROTEIN.  
 GN CI-D.  
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
 CC EUKARYOTA: METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OREGON-R;  
 RX MEDLINE; 90346286.  
 RA ORENIC T.V., SLESARSKI D.C., KROLL K.L., HOLMGREN R.A.;  
 RL GENES DEV. 4:1053-1067(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OREGON-R;  
 RX MEDLINE; 92146935.  
 RA BERRY A.J., AJIOKA J.W., KREITMAN M.;  
 RL GENETICS 129:1111-1117(1991).  
 CC -1- FUNCTION: INVOLVED IN SEGMENT POLARITY. IS REQUIRED FOR THE NORMAL  
 CC DEVELOPMENT OF THE POSTERIOR HALF OF EACH EMBRYONIC SEGMENT.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED UNIFORMLY THROUGHOUT THE BLASTODERM  
 CC STAGE AND GASTRULATION AND DOES NOT RESOLVE INTO SEGMENTALLY  
 CC REPEATING STRIPES UNTIL THE END OF THE SHORT PHASE OF GERM-BAND  
 CC EXTENSION.  
 CC -1- SIMILARITY: TO THE GLI-RELATED GROUP OF C2H2-TYPE ZINC-FINGERS  
 CC PROTEINS.  
 DR EMBL; X54360; G7733;  
 DR PIR; A35817; A35817;  
 DR PIR; S12769; S12769.  
 DR HSSP; P07248; IARD.  
 DR FLYBASE; FBgn0004859; c1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_4.  
 KM DEVELOPMENTAL PROTEIN; SEGMENTATION POLARITY PROTEIN; ZINC-FINGER;  
 KM METAL-BINDING; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.  
 FT DOMAIN 451 603 ZINC-FINGERS.  
 FT ZN\_FING 451 476 C2H2-TYPE.

FT ZN\_FING 484 511 C2H2-TYPE.  
 FT ZN\_FING 517 541 C2H2-TYPE.  
 FT ZN\_FING 547 572 C2H2-TYPE.  
 FT ZN\_FING 578 603 C2H2-TYPE.  
 SQ SEQUENCE 1377 AA: 150881 MW; A14EB3FC CRC32;

Query Match  
 Best Local Similarity 59.6%; Score 53; DB 1; Length 1377;  
 Matches: 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 519 CTFECRFA 527  
 OY 1 CILESCFRA 9

RESULT 13  
 ID VAA-SOLIN STANDARD; PRT: 117 AA.

AC P35777;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE VENOM ALLERGEN IV (ALLERGEN SOL I 4) (SOL I IV).  
 OS SOLENOPODIS INVICTA (RED IMPORTED FIRE ANT).  
 OC ENKARYOTA; METAZOA; ARTHROPODA; INSECTA; HYMENOPTERA.  
 RN [1]  
 RP SEQUENCE.

RC TISSUE-VENOM;  
 RX MEDLINE: 93139387.  
 RA HOFFMAN D.R.;  
 RL J. ALLERGY CLIN. IMMUNOL. 91:71-78(1993).  
 RN [2]  
 RP PARTIAL SEQUENCE OF 1-31.  
 RC TISSUE-VENOM;  
 RX MEDLINE: 90285439.  
 RA HOFFMAN D.R., SMITH A.M., SCHMIDT M., MOFFITT J.E., GURALNICK M.,  
 RL J. ALLERGY CLIN. IMMUNOL. 85:988-996(1990).  
 CC -1- DISEASE: THE MOST COMMON CAUSE OF INSECT VENOM ALLERGY IN THE  
 CC SOUTHEASTERN UNITED STATES IS THE IMPORTED FIRE ANT.  
 CC -1- SIMILARITY: MONOMER.  
 CC -1- SIMILARITY: TO VENOM ALLERGEN II.  
 DR PIR: C37330; C37330.  
 KW VENOM; ALLERGEN.  
 FT VARIANT 23 R -> H.  
 FT VARIANT 37 L -> I.  
 SQ SEQUENCE 117 AA: 13341 MW; 7174AB01 CRC32;

Query Match  
 Best Local Similarity 58.4%; Score 52; DB 1; Length 117;  
 Matches: 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 12 ILEKICRTV 20  
 OY 2 ILESCFRAV 10

RESULT 14  
 ID YIOL YEAST STANDARD; PRT: 129 AA.

AC P40461;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 14.4 KD PROTEIN IN COT2-AXL2 INTERGENIC REGION.  
 GN Y1141W  
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,  
 RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,  
 RA GENIES S., HAWLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,  
 RA LOUIS E., IYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,

RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,  
 RA WALSH S.V., WHITEHEAD S.;  
 RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: Z47047; G763205; -;  
 DR EMBL: Z38059; G557782; -;  
 DR PIR: S48393; S48393.  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 129 AA: 14437 MW; 48F132D8 CRC32;

Query Match  
 Best Local Similarity 58.4%; Score 52; DB 1; Length 129;  
 Matches: 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 7 VLEPCKNVI 16  
 OY 2 ILESCFRAV 11

RESULT 15  
 ID CBPL CANAL STANDARD; PRT: 489 AA.

AC P31225;  
 DT 01-JUL-1993 (REL. 26, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE CORTICOSTEROID-BINDING PROTEIN.  
 GN CBPL.  
 OS CANDIDA ALBICANS (YEAST)  
 OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 53-63; 373-380 AND 403-414.  
 RC STRAIN-STN-1;  
 RX MEDLINE: 93189605.  
 RA MALLORY P.J., ZHAO X., MADANI N.D., FELDMAN D.,  
 RL PROC. NATL. ACAD. SCI. U.S.A. 90:1902-1906(1993).  
 CC -1- FUNCTION: MAY BE A FLAVOPROTEIN WITH ENZYMIC ACTIVITY.  
 CC -1- SIMILARITY: TO YEAST FMS1.  
 DR EMBL: L08824; E77215; -;  
 DR PIR: A47259; A47259.  
 KW STEROID-BINDING.  
 SQ SEQUENCE 489 AA: 55492 MW; CC69C62A CRC32;

Query Match  
 Best Local Similarity 58.4%; Score 52; DB 1; Length 489;  
 Matches: 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 472 CILENFRNDV 482  
 OY 1 CILESCFRAV 11

Search completed: Fri Sep 11 12:47:02 1998  
 Job time: 7 secs.

(W.T.)

Distribution rights by Oxford Molecular Ltd

t generated.

1 CILIESCFRAVI 11

## Sap 15

Twisting first 45

13:sp\_unclassified

ved by analysis of the total score distribu

## SUMMARIES

3.6 362 3 016337

|    |    |      |      |   |        |
|----|----|------|------|---|--------|
| 45 | 51 | 57.3 | 2700 | 3 | Q19689 |
|----|----|------|------|---|--------|

## ALIGNMENTS

| AD                    | 016890.  | PELLEGRINIARI                          | FRI           | 232 AA.    |
|-----------------------|--|--|---------------|------------|
| AD                    | 016890.  |  |               |            |
| DT                    | 01-JAN-1998  | (TREMBLEL. 05, CREATED)                |               |            |
| DT                    | 01-JAN-1998  | (TREMBLEL. 05, LAST SEQUENCE UPDATE)   |               |            |
| DT                    | 01-JAN-1998  | (TREMBLEL. 05, LAST ANNOTATION UPDATE) |               |            |
| DE                    | F13A2.8  | PROTEIN.                               |               |            |
| CN                    | F13A2.8.   |  |               |            |
| OS                    | CAENORHABDITIS ELEGANS.                                    |  |               |            |
| OC                    | EUTARICOTA; METAZOA; ACCELLOMATES; NEMATODA; SECCERNENTEA; |  |               |            |
| KN                    | [1]  |  |               |            |
| RP                    | SEQUENCE FROM N.A.   |  |               |            |
| RC                    | STRAIN-BRISTOL N2.   |  |               |            |
| RC                    | MEDLINE: 94150718.   |  |               |            |
| RA                    | WILSON R., AINSOUGH R., CONNELL M., COPSEY T., COOPER J.   |  |               |            |
| RA                    | BOWFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.   |  |               |            |
| RA                    | CRAXTON M., DEAR S., DU Z., DUBBIN R., FAVELLO A., FUTRO   |  |               |            |
| RA                    | GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., J   |  |               |            |
| RA                    | JONES M., KERSHAN J., KIRSTEN J., LAISTER N., LATREILLE I  |  |               |            |
| RA                    | LIGHTNING J., LOYD C., MCGURRAY A., MORTIMORE B., O'CALL   |  |               |            |
| RA                    | PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SANDERS D.,    |  |               |            |
| RA                    | SALMON D.N., SMITH A., SONNHAMER E., STADEN R., STUSTON    |  |               |            |
| RA                    | THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATER   |  |               |            |
| RA                    | WATSON A., WENUSOCK L., WILKINSON-SPROAT J., WOHLDMAN P    |  |               |            |
| RL                    | Nature 368:32-38(1994).                                    |  |               |            |
| KN                    | [2]  |  |               |            |
| RP                    | SEQUENCE FROM N.A.   |  |               |            |
| RC                    | STRAIN-BRISTOL N2.   |  |               |            |
| RA                    | MCPHERSON J., BRADSHAW H.;                                 |  |               |            |
| RL                    | SUBMITTED (SEP-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.      |  |               |            |
| KN                    | [3]  |  |               |            |
| RP                    | SEQUENCE FROM N.A.   |  |               |            |
| RC                    | STRAIN-BRISTOL N2.   |  |               |            |
| RA                    | WATERSTON R.;  |  |               |            |
| RL                    | SUBMITTED (SEP-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.      |  |               |            |
| CC                    | -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).         |  |               |            |
| DR                    | EMBL; AF022970; G2384801; -                                |  |               |            |
| DR                    | PROSITE: PS00031, NUCLEAR RECEPTOR; 1.                     |  |               |            |
| KW                    | RECEPTOR; TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR   |  |               |            |
| KW                    | ZINC-FINGER.   |  |               |            |
| SO                    | SEQUENCE 232 AA: 26943 MW: 84973012 CRC32;                 |  |               |            |
| Query Match           | 67.4%;   | Score 60:                              | DB 3:         | Length 232 |
| Best Local Similarity | 77.8%;   | Pred.                                  | No. 1.87e-01; |            |

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 72 CRLKCFRA 80  
Matches 1; Conservative 9

RESULT 2  
ID 049624; PRELIMINARY; PRT: 2118 AA.

AC 049624;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DE 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)

DE PROBABLE MYCOCEROSIS ACID SYNTHASE.

GN MASA OR B1170 C2 209.

OS MYCOBACTERIUM LEPRAE.

OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RA ROBINSON K., SMITH D.R.;

RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH METHYLMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO FORM MYCOCEROSYL LIPIDS.

CC -1- COFACTOR: CONTAINS ONE COVALENTLY BOUND PHOSPHOPANTHETHEINE.

CC -1- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL ARRANGEMENT.

CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.

CC -1- SIMILARITY: PARTIAL TO S.ERYTHRAEA ERYTHRONOLIDE SYNTHASE, MODULE 4, AND TO VERTEBRATE FATTY ACID SYNTHASES.

DR EMBL; U00010; G466793; -

KW HYPOTHETICAL PROTEIN; FATTY ACID BIOSYNTHESIS; MULTIFUNCTIONAL ENZYME; PHOSPHOPANTHETHEINE; TRANSFERASE; HYDROLASE; OXIDOREDUCTASE; LIGASE;

KM NADP; MEMBRANE.

FT DOMAIN 1 ? ? BETA-KETOACYL SYNTHASE.

FT DOMAIN 2 ? ? ACYL TRANSFERASE.

FT DOMAIN 3 ? ? ENOYL REDUCTASE.

FT DOMAIN 4 ? ? BETA-KETOACYL REDUCTASE.

FT ACT\_SITE 178 178 ACYL CARRIER.

FT ACT\_SITE 624 624 BETA-KETOACYL SYNTHASE (BY SIMILARITY).

FT NP\_BIND 1567 1584 NADP (KR) (BY SIMILARITY).

FT BINDING 1771 1786 NADP (ER) (BY SIMILARITY).

FT BINDING 2069 2069 PHOSPHOPANTHETHEINE (BY SIMILARITY).

SO SEQUENCE 2118 AA; 226497 MW; 764905A4 CRC32;

Query Match

Best Local Similarity 65.28; Score 58; DB 9; Length 2118;

Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 1099 LUDACFOSVI 1108

QY 2 ILSCFRAVI 11

RESULT 3

ID P89938; PRELIMINARY; PRT: 1727 AA.

AC P89938;

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)

DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)

DE 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)

DE REPLICASE ORF1A POLYPROTEIN.

OS EQUINE ARTERITIS VIRUS (EAV)

OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; TOGAVIRIDAE; ARTERIVIRUSES.

RN [1]

RP SEQUENCE FROM N.A.

RA VAN DINTEN L.C., DEN BOON J.A., MASSENAAR A.L.M., SPAN W.J.M.,

RL PROC. NATL. ACAD. SCI. U.S.A. 94:991-996(1997).

DR EMBL; Y07862; E280813; -

KW POLYPROTEIN.

SO SEQUENCE 1727 AA; 186986 MW; A602D83D CRC32;

Query Match

Best Local Similarity 64.08; Score 57; DB 11; Length 1727;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 344 CLDESCFRGI 353

QY 1 CLDESCFRAY 10

RESULT 4

ID P89939; PRELIMINARY; PRT: 3175 AA.

AC P89939;

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)

DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)

DE 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)

DE REPLICASE ORF1B POLYPROTEIN.

OS EQUINE ARTERITIS VIRUS (EAV).

OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; TOGAVIRIDAE; ARTERIVIRUSES.

RN [1]

RP SEQUENCE FROM N.A.

RA VAN DINTEN L.C., DEN BOON J.A., MASSENAAR A.L.M., SPAN W.J.M.,

RL PROC. NATL. ACAD. SCI. U.S.A. 94:991-996(1997).

DR EMBL; Y07862; E280872; -

KW POLYPROTEIN.

SO SEQUENCE 3175 AA; 345374 MW; FDFD6351 CRC32;

Query Match

Best Local Similarity 64.08; Score 57; DB 11; Length 3175;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 344 CLDESCFRGI 353

QY 1 CLDESCFRAY 10

RESULT 5

ID Q18174; PRELIMINARY; PRT: 880 AA.

AC Q18174;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL PROTEIN C25G4.10 (FRAGMENT).

GN C25G4.10.

OS CAENORHABDITIS ELEGANS.

OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.

RN [1]

RP SEQUENCE FROM N.A.

RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,

RA COULSON A., CRAXTON M., DEAR S., DU Z., DUREIN R., FAVELLO A.,

RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,

RA JOHNSON L., JONES M., KERSHAW J., KRISTEN J., LAISTER N.,

RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,

RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFKEN L., ROOPRA A.,

RA SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONNHAMMER E.,

RA STADEN R., SUSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,

RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,

RA WILKINSON-SPROAT J., WOHLDMAN P.,

RL NATURE 368:32-38(1994).

DR EMBL; 270680; E1187502; -

KW HYPOTHETICAL PROTEIN.

FT NON\_TER 880 880

SO SEQUENCE 880 AA; 98049 MW; BC8C632C CRC32;

Query Match

Best Local Similarity 62.98; Score 56; DB 3; Length 880;

Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 61 CVALNCSKDI 71



OY 1 CILESCFRAVI 11

RESULT 6 PRELIMINARY; PRT; 86 AA.

AC 004820;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DE 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 9.1 KD PROTEIN.  
OC EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; VIRIDIPLANTAE;  
OC CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA; MAGNOLIOPHYTA;  
OC LILIOPSIDA; POALES; POACEAE; SPOBOLOUS.

RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LEAF;  
RA BLOMSTEDT C.K., GIANELLO R.D., NEALE A.D., HAMILL J.D., GAFF D.F.;  
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Y10784; E294072; -.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 86 AA; 9148 MW; CB6C0196 CRC32;

Query Match 60.7%; Score 54; DB 8; Length 86;  
Best Local Similarity 57.1%; Pred. No. 3.19e+00;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 80 CILDACF 86  
OY 1 CILESCF 7

RESULT 7 PRELIMINARY; PRT; 200 AA.

AC 041194;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE ENVELOPE PROTEIN.  
GN ENV.  
OS PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS.  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; ARTERIVIRUS.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NADC-9 (1A);  
RA ANDREYEV V.G., WESLEY R.D., MENGELING W.L., VORWALD A.C., LAGER K.M.;  
RL ARCH. VIROL. 142:993-1001(1997).  
DR EMBL: U66393; G2231279; -.  
KW ENVELOPE PROTEIN.

SQ SEQUENCE 200 AA; 22248 MW; B7CE6C92 CRC32;

Query Match 60.7%; Score 54; DB 11; Length 200;  
Best Local Similarity 54.5%; Pred. No. 3.19e+00;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 19 C1VSCFVALV 29  
OY 1 C1VSCFRAVI 11

RESULT 8 PRELIMINARY; PRT; 647 AA.

AC 027790;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE DNA MISMATCH RECOGNITION PROTEIN MUTS.  
GN MTH1762.  
OS METHANOBACTERIUM THERMOAUTOTROPHICUM.  
OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOBACTERIALES;  
OC METHANOBACTERIACEAE.  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-DELTA H;  
RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,  
RA ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,  
RA HARRISON D., HOANG L., KEAGLE P., LOMM W., POTTER B., QIU D.,  
RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,  
RA JIMANI N., CAROSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,  
RA MCDONALD S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,  
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;  
RL J. BACTERIOL. 179:7135-7155(1997).

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DELTA H;  
RA SMITH D.R.;  
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AE000931; G2622891; -.  
SQ SEQUENCE 647 AA; 73592 MW; 499DE792 CRC32;

Query Match 60.7%; Score 54; DB 9; Length 647;  
Best Local Similarity 60.0%; Pred. No. 3.19e+00;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 487 CALEACRVV 496  
OY 1 C1VSCFRAVI 10

RESULT 9 PRELIMINARY; PRT; 200 AA.

AC 041195;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE ENVELOPE PROTEIN.  
GN ENV.  
OS PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS.  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; ARTERIVIRUS.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NADC-8 (1A);  
RA ANDREYEV V.G., WESLEY R.D., MENGELING W.L., VORWALD A.C., LAGER K.M.;  
RL ARCH. VIROL. 142:993-1001(1997).  
DR EMBL: U66394; G2231281; -.  
KW ENVELOPE PROTEIN.

SQ SEQUENCE 200 AA; 22261 MW; 8FBD7B93 CRC32;

Query Match 59.6%; Score 53; DB 11; Length 200;  
Best Local Similarity 54.5%; Pred. No. 5.04e+00;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 19 C1VSCFVALV 29  
OY 1 C1VSCFRAVI 11

RESULT 10 PRELIMINARY; PRT; 200 AA.

AC 041192;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE ENVELOPE PROTEIN.  
GN ENV.  
OS PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS.  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; ARTERIVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-30352-3 (H1);  
RA ANDREYEV V.G., WESLEY R.D., MENGELING W.L., VORWALD A.C., LAGER K.M.;  
RL ARCH. VIROL. 142:993-1001(1997).  
DR EMBL: U66391; G2231275; -.  
KW ENVELOPE PROTEIN.  
SQ SEQUENCE 200 AA; 22228 MW; 3861127F CRC32;

Query Match  
Best Local Similarity 54.5%; Score 53; DB 11; Length 200;  
Pred. No. 5.04e+00;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 19 CIVPSCFVALV 29  
11:111111:  
1 CILSCFRAVI 11

RESULT 11  
ID 041198 PRELIMINARY; PRT; 200 AA.  
AC 041199;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE ENVELOPE PROTEIN.  
GN ENV.  
OS PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS.  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; ARTERIVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-46907 (KY);  
RA ANDREYEV V.G., WESLEY R.D., MENGELING W.L., VORMALD A.C., LAGER K.M.;  
RL ARCH. VIROL. 142:993-1001(1997).  
DR EMBL; U66398; G2231289; -.  
KW ENVELOPE PROTEIN.  
SQ SEQUENCE 200 AA; 22304 MW; FA431E37 CRC32;

Query Match  
Best Local Similarity 54.5%; Score 53; DB 11; Length 200;  
Pred. No. 5.04e+00;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 19 CIVPSCFVALV 29  
11:111111:  
1 CILSCFRAVI 11

RESULT 12  
ID 041193 PRELIMINARY; PRT; 200 AA.  
AC 041193;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE ENVELOPE PROTEIN.  
GN ENV.  
OS PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS.  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; ARTERIVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-5556 (MT);  
RA ANDREYEV V.G., WESLEY R.D., MENGELING W.L., VORMALD A.C., LAGER K.M.;  
RL ARCH. VIROL. 142:993-1001(1997).  
DR EMBL; U66392; G2231277; -.  
KW ENVELOPE PROTEIN.  
SQ SEQUENCE 200 AA; 22281 MW; E51F9C9B CRC32;

Query Match  
Best Local Similarity 54.5%; Score 53; DB 11; Length 200;  
Pred. No. 5.04e+00;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 19 CIVPSCFVALV 29  
11:111111:  
1 CILSCFRAVI 11

RESULT 13  
ID 041198 PRELIMINARY; PRT; 200 AA.  
AC 041198;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE ENVELOPE PROTEIN.

GN ENV.  
OS PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS.  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; ARTERIVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-1205-D (MO);  
RA ANDREYEV V.G., WESLEY R.D., MENGELING W.L., VORMALD A.C., LAGER K.M.;  
RL ARCH. VIROL. 142:993-1001(1997).  
DR EMBL; U66397; G2231287; -.  
KW ENVELOPE PROTEIN.  
SQ SEQUENCE 200 AA; 22366 MW; 5BB77B7F CRC32;

Query Match  
Best Local Similarity 54.5%; Score 53; DB 11; Length 200;  
Pred. No. 5.04e+00;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 19 CIVPSCFVALV 29  
11:111111:  
1 CILSCFRAVI 11

RESULT 14  
ID 041197 PRELIMINARY; PRT; 200 AA.  
AC 041197;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE ENVELOPE PROTEIN.  
GN ENV.  
OS PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS.  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; ARTERIVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-46448-12 (IA);  
RA ANDREYEV V.G., WESLEY R.D., MENGELING W.L., VORMALD A.C., LAGER K.M.;  
RL ARCH. VIROL. 142:993-1001(1997).  
DR EMBL; U66396; G2231285; -.  
KW ENVELOPE PROTEIN.  
SQ SEQUENCE 200 AA; 22304 MW; FDA3048A CRC32;

Query Match  
Best Local Similarity 54.5%; Score 53; DB 11; Length 200;  
Pred. No. 5.04e+00;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 19 CIVPSCFVALV 29  
11:111111:  
1 CILSCFRAVI 11

RESULT 15  
ID 084936 PRELIMINARY; PRT; 200 AA.  
AC 084936;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE GICOSYLATED.  
GN PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS.  
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; TOGAVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94282932.  
RA MENG X.J., PAUL P.S., HALBUR P.G.;  
RL J. GEN. VIROL. 75:1795-1801(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93174942.  
RA CONZELMANN K.K., VISSER N., VAN WOENSEL P., THIEL H.J.;  
RL VIROLOGY 193:329-339(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93297139.  
RA MEIJENBERG J.J., HULST M.M., DE MEIJER E.J., MOONEN P.L.,  
RA BESTEN A., DE KLUYVER E.P., WENVOORT G., MOORMANN R.J.;

Sun Sep 13 10:56:38 1998

US-08-452-843-3.rspt

Page 5

RL VIROLOGY 192:62-72(1993).  
DR EMBL: 003040; G515975;  
SQ SEQUENCE 200 AA; 22201 MW; F75998BD CRC32;

Query Match 59.68; Score 53; DB 11; Length 200;  
Best Local Similarity 54.58; Pred. No. 5.04e+00;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 19 CIVPSCFVALV 29  
11:11111:  
QY 1 CIVPSCFRAVI 11

Search completed: Fri Sep 11 12:47:48 1998  
Job time : 27 secs.

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on;      Fri Sep 11 12:41:36 1998;  MasPar time 2.60 Seconds
```

Tabular output not generated.

|              |                           |
|--------------|---------------------------|
| Title:       | >US-08-452-843-2          |
| Description: | (1-9) from US08452843.pep |

Sequence: 1 YPKVKQWPL

Scoring table: PAM 150

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum 'Match' 0%

## Listing first 45 summaries

Database:

a-geneseq32

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 18.000; Variance 53.697; scale 0.3335

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description             | Pred. No. |
|------------|-------|-------------|--------|----|--------|-------------------------|-----------|
| 1          | 82.1  | 100.0       | 9      | 18 | R89363 | Immunogenic peptide,    | 2.55e-02  |
| 2          | 70    | 85.4        | 9      | 15 | R78851 | HIV pol 185-193 cytot   | 6.65e-01  |
| 3          | 70    | 85.4        | 9      | 14 | R70601 | HIV(HB3)POL-9, human    | 6.65e-01  |
| 4          | 70    | 85.4        | 25     | 25 | W32893 | HIV pol protein epitope | 6.65e-01  |
| 5          | 70    | 85.4        | 25     | 12 | P68755 | Cytotoxic T lymphocyte  | 6.65e-01  |
| 6          | 70    | 85.4        | 49     | 16 | R94761 | CTL epitopes pol1 der   | 6.65e-01  |
| 7          | 70    | 85.4        | 53     | 2  | R29709 | p41 gag protein from    | 6.65e-01  |
| 8          | 70    | 85.4        | 54     | 3  | P61510 | Sequence of pol prote   | 6.65e-01  |
| 9          | 70    | 85.4        | 91     | 2  | R08053 | AcNPV-HIVR-pol prote    | 6.65e-01  |
| 10         | 70    | 85.4        | 91     | 2  | R08057 | AcNPV-HIVR-pol prote    | 6.65e-01  |
| 11         | 70    | 85.4        | 96     | 2  | R03501 | HIV-1 pol protein of    | 6.65e-01  |
| 12         | 70    | 85.4        | 100    | 3  | R12256 | Sequence deduced from   | 6.65e-01  |
| 13         | 70    | 85.4        | 100    | 3  | P61508 | HIV-1 strain OYI POL    | 6.65e-01  |
| 14         | 70    | 85.4        | 100    | 3  | R08061 | Sequence of ARV-2 (B    | 6.65e-01  |
| 15         | 70    | 85.4        | 100    | 3  | P60420 | HIV-1 pol protein of    | 6.65e-01  |
| 16         | 70    | 85.4        | 100    | 3  | R08059 | Sequence of LAV virus   | 6.65e-01  |
| 17         | 70    | 85.4        | 100    | 6  | R29705 | HIV-1 pol protein of    | 6.65e-01  |
| 18         | 70    | 85.4        | 100    | 2  | P70861 | pol gene deduced from   | 6.65e-01  |
|            |       |             |        |    |        | Sequence encoded by L   | 6.65e-01  |

|    |    |      |      |    |        |                       |          |
|----|----|------|------|----|--------|-----------------------|----------|
| 19 | 70 | 85.4 | 1004 | 2  | R08058 | HIV-1 pol protein of  | 6.55e-01 |
| 20 | 70 | 85.4 | 1010 | 19 | R19123 | Human immunodeficienc | 6.55e-01 |
| 21 | 70 | 85.4 | 1012 | 3  | P61507 | Sequence of reverse t | 6.55e-01 |
| 22 | 70 | 85.4 | 1015 | 8  | R43875 | HTLV-II pol gene pro  | 6.55e-01 |
| 23 | 70 | 85.4 | 1015 | 8  | R43867 | HTLV-II pol gene pro  | 6.55e-01 |
| 24 | 70 | 85.4 | 1015 | 3  | P60347 | HTLV-II virus (HIV v  | 6.55e-01 |
| 25 | 70 | 85.4 | 1016 | 2  | R08063 | HIV-1 pol protein of  | 6.55e-01 |
| 26 | 70 | 85.4 | 1016 | 2  | R08052 | ACNPV-HIVp01 protei   | 6.55e-01 |
| 27 | 70 | 85.4 | 1016 | 2  | R08054 | HIV-1 pol protein of  | 6.55e-01 |
| 28 | 70 | 85.4 | 1022 | 1  | P81854 | Sequence encoded by L | 6.55e-01 |
| 29 | 70 | 85.4 | 1491 | 1  | P91048 | Transcription sequenc | 6.55e-01 |
| 30 | 70 | 85.4 | 2033 | 2  | R08056 | HIV-1 pol protein of  | 6.55e-01 |
| 31 | 70 | 85.4 | 2033 | 3  | R08055 | HIV-1 pol protein of  | 6.55e-01 |
| 32 | 67 | 81.7 | 1002 | 1  | P81861 | Sequence encoded by L | 1.47e+00 |
| 33 | 67 | 81.7 | 1003 | 2  | R08060 | HIV-1 pol protein of  | 1.47e+00 |
| 34 | 67 | 81.7 | 1055 | 24 | W13055 | HIV-2 provirus-encode | 1.47e+00 |
| 35 | 65 | 79.3 | 1009 | 2  | P10275 | Simian immunodefici   | 2.49e+00 |
| 36 | 64 | 78.0 | 1056 | 1  | P81783 | Sequence encoded by p | 3.24e+00 |
| 37 | 64 | 78.0 | 1056 | 1  | R08089 | Sequence of pol prote | 3.24e+00 |
| 38 | 63 | 76.8 | 1060 | 4  | R23365 | Sequence of pol prote | 3.24e+00 |
| 39 | 63 | 76.8 | 1124 | 2  | R24237 | SIVmac339 pol gene pr | 3.24e+00 |
| 40 | 61 | 74.4 | 230  | 12 | R6767  | pol polypeptide of fi | 4.20e+00 |
| 41 | 61 | 74.4 | 1014 | 1  | P80810 | HIV-2 ROD protease re | 7.04e+00 |
| 42 | 61 | 74.4 | 1027 | 1  | P81773 | Sequence of pol prote | 7.04e+00 |
| 43 | 61 | 74.4 | 1036 | 4  | R20599 | Sequence encoded by o | 7.04e+00 |
| 44 | 60 | 73.2 | 1053 | 1  | R05614 | ROD HIV-2 polymerase. | 7.04e+00 |
| 45 | 58 | 70.7 | 1035 | 2  | R04025 | RIV pol gene product. | 9.10e+00 |
|    |    |      |      |    |        | pol gene product of c | 1.52e+01 |

## ALIGNMENTS

RESULT 1

ID R89363 standard; peptide; 9 AA.

AC R89363:

DT 18-SEP-1996 (first entry).

DE Immunogenic peptide, based on Y1 analog of 1054.05.

KW Immunogenic peptide; supermotif; HLA molecule; CTL response;

RW therapeutic; diagnostic; cancer; viral infection; hepatitis B;

KW hepatitis C.

OS Synthetic.

PN MO6903140-AL.

PD 08-FEB-1996.

FE 21-JUL-1995; 009234.

PR 21-JUL-1994; US-278634.

PR 23-NOV-1994; US-344824.

PR 30-MAY-1995; US-452843.

PA (CYTE-) CYTEL CORP.

PI Sette A; Sidney J;

DR WPI; 96-116784/12.

FT Compr. comprising immunogenic peptide with supermotif allowing more

PT than one HLA mol. to bind - used to induce CTL response in patient

PT and for in vivo and ex vivo therapeutic and diagnostic applications

PS Claim 2; Page 26; 32pp; English.

CC The sequences given in R89362-82 are immunogenic peptides which were

CC used in the composition of the invention. The composition comprises

CC an immunogenic peptide of 9-10 residues with a supermotif which

CC allows binding of more than one HLA molecule. It pref. comprises

CC two conserved residues, a first at the 2nd position from the N-

CC terminal is Pto, and a 2nd at the C-terminal is Met. These peptides

CC are used to induce a CTL response in a patient. They are also

CC useful in compositions for in vivo and ex vivo therapeutic and

CC diagnostic applications, e.g the treatment of cancer and viral

CC infections, e.g. hepatitis B and C.

CC Sequence 9 AA;

CC

|                       |         |                     |        |                                |
|-----------------------|---------|---------------------|--------|--------------------------------|
| Query Match           | 100.0%; | Score 82;           | DB 18; | Length 9;                      |
| Best Local Similarity | 100.0%; | Pred. No. 2.55e-02; |        |                                |
| Matches               | 9;      | Conservative        | 0;     | Mismatches 0; Indels 0; Gaps 0 |

**Dd**    1 ypkvkqwpL 9  
         |||||  
**Qy**    1 ypkvkQwPL 9

RESULT 2  
ID R78851 standard; peptide; 9 AA.  
AC R78851;  
DT 27-MAR-1996 (first entry)  
DE HIV pol 185-193 cytotoxic T lymphocyte epitope.  
KW HIV pol 185-193; cytotoxic T; CTL; epitope; helper T; HTL; cell;  
KW lymphocyte; viruses; parasites; tumours; antigens; treatment;  
KW disease prevention.  
OS Human immunodeficiency virus.  
PN WO9522317-A1.  
PD 24-AUG-1995; 11.  
PR 16-FEB-1996; US-02121.  
PR 16-FEB-1994; US-197484.  
PA (CYTEC) CYTEL CORP.  
PI Ceut RW, Grey H, Sette AD, Vitello MA;  
DR WPI; 95-302545/39.  
PT Compn. inducing cytotoxic T lymphocyte response to pref. viral,  
PT bacterial, parasitic or tumour antigens - useful in the treatment  
PT and prevention of diseases associated with the antigen e.g.  
PT hepatitis B.  
PS Disclosure; Page 17; 109pp; English.  
CC A compn. which induces a cytotoxic T lymphocyte (CTL) response to  
CC an antigen (Ag) in a mammal comprises, a CTL Ag response inducing  
CC peptide (i.e. R78824-R78853) and a lipid conjugated helper T cell  
CC inducing peptide. The compn. induces a CTL response to bacterial,  
CC viral or tumour Ags, and is therefore useful in the treatment and  
CC prevention of diseases associated with the Ag.  
SQ Sequence 9 AA:

Query Match 85.4%; Score 70; DB 15; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.65e-01;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 2 PKVKQWPL 9  
OY 2 PKVKQWPL 9

RESULT 3  
ID R70601 standard; Peptide; 9 AA.  
AC R70601;  
DT 14-FEB-1996 (first entry)  
DE HIV(835)POL-9, human immunodeficiency virus epitope.  
KW HLA; human lymphocyte antigen; HIV; human immunodeficiency virus;  
KW binding peptide; induce killer cell; prevention; treatment; AIDS;  
KW auto-immune disease syndrome; vaccine.  
OS Human immunodeficiency virus.  
PN WO9511255-A1.  
PD 27-APR-1995.  
PR 19-OCT-1994; J01756.  
PR 19-OCT-1993; JP-261302.  
PA (AJIN) AJINOMOTO CO INC.  
PA (AJIN) AJINOMOTO KK.  
PI Miwa K, Takiguchi M;  
DR WPI; 95-170188/22.  
PT HLA-binding peptide fragments from HIV proteins - induce killer  
PT cells which target HIV-infected cells and can be incorporated into  
PT anti-HIV vaccines.  
PS Example 1; Page 10; 61pp; Japanese.  
CC R70601 is a peptide fragment derived from an HIV (Human Immunodeficiency  
CC virus) protein and is capable of binding to a human lymphocyte antigen.  
CC The peptide can induce killer cells which target HIV-infected cells.  
CC It is also useful in the prevention and treatment of HIV and AIDS.  
CC Anti-HIV vaccines may incorporate the peptides, or may incorporate a  
CC vector (such as vaccinia or BCG) contg. DNA encoding the peptides.  
SQ Sequence 9 AA:

Query Match 85.4%; Score 70; DB 14; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.65e-01;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 2 PKVKQWPL 9

OY 2 PKVKQWPL 9

RESULT 4  
ID W32893 standard; peptide; 25 AA.  
AC W32893;  
DT 16-JAN-1998 (first entry)  
DE HIV pol protein epitope 40.  
KW Hydrophilic; antigenic determinant; HIV; envelope; glycoprotein;  
KW env; gp; recognition; B lymphocyte; type specific; antibody;  
KW vaccine; protection; immune response; infection; neutralisation;  
KW epitope.  
OS Human immunodeficiency virus.  
PN WO9714436-A1.  
PD 24-APR-1997.  
PR 18-OCT-1996; U16911.  
PR 09-FEB-1996; US-599266.  
PR 20-OCT-1995; US-546515.  
PA (UYDU-) UNIV DUKE.  
PI Haynes BF, Parker TJ;  
DR WPI; 97-244862/22.  
PT Synthetic human immunodeficiency virus vaccine - comprising  
PT hydrophilic peptide corresponding to at least 1 antigenic  
PT determinant of envelope glyco:protein recognised by B lymphocytes  
PS Disclosure; Page 27; 104pp; English.  
CC An essentially pure hydrophilic peptide, comprising at least 1  
CC antigenic determinant of human immunodeficiency virus (HIV)  
CC envelope (env) glycoprotein (gp) recognised by B lymphocytes,  
CC when covalently linked to a carrier molecule, i.e. the present  
CC sequence, induces the production of high titres of protective, type  
CC specific anti-HIV antibodies (Ab) in a mammal. The peptide can be  
CC used in vaccines for producing a protective immune response to HIV  
CC infection, while a HIV neutralising Ab can be induced in a primate  
CC by administering a composition comprising HIV env peptides that  
CC disrupt gp120/gp41 interactions.  
SQ Sequence 25 AA:

Query Match 85.4%; Score 70; DB 25; Length 25;  
Best Local Similarity 100.0%; Pred. No. 6.65e-01;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 15 PKVKQWPL 22  
OY 2 PKVKQWPL 9

RESULT 5  
ID R68755 standard; peptide; 25 AA.  
AC R68755;  
DT 23-AUG-1995 (first entry)  
DE Cytotoxic T lymphocyte epitope 12 derived from pol protein.  
KW cytotoxic T lymphocyte; epitope; antigen; pathogenic; nef; gag; pol;  
KW env; gp120; gp41; HIV; cell-mediated immunity;  
KW human immunodeficiency virus; class I restricted.  
OS Human immunodeficiency virus.  
PN WO9428871-A.  
PD 22-DEC-1994.  
PR 07-JUN-1994; U06394.  
PR 07-JUN-1993; US-072718.  
PA (ENDO-) ENDOCON INC.  
PI Leonard RJ;  
DR WPI; 95-036067/05.  
PT Implant for sustained release of pathogen-associated antigen -  
PT forming chronic inflammatory site producing cytotoxic  
PT T-lymphocytes lysing infected cells, esp. for treating AIDS  
PS Disclosure; Page 11; 35pp; English.  
CC R68744-805 are cytotoxic T lymphocyte (CTL) class I and II restricted  
CC epitopes derived from human immunodeficiency virus proteins. R68755  
CC corresponds to amino acid residues 172-196 of the pol protein. These  
CC antigens are examples of peptides that can be used with an immunogenic  
CC implant. The implant is associated with an antigen associated with a  
CC pathogen and used to form a discrete, localised chronic inflammation

CC site which acts as a local 'factory' for prodn. of CTL's which lyse  
 CC cells infected with a specific pathogen. The expanded set of  
 CC pathogen-specific CTL's can eradicate or prevent development of  
 CC infection, and can also be used to treat or arrest the development of  
 CC cancers associated with infection.  
 SQ Sequence 25 AA;

Query Match 85.4%; Score 70; DB 12; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 6.65e-01;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DE 9.15 PKVKQWPL 22  
 |||||  
 QY 2 PKVKQWPL 9

RESULT 6  
 ID R94761 standard; Protein; 49 AA.  
 AC R94761;  
 DT 11-JUN-1996 (first entry)  
 DE CTL epitopes pool derived from PCPOLT5A.  
 KW Canarypox; CPV; ALVAC; attenuated; therapy; prevention; rabies;  
 KM vector; vaccine; antibody; CTL1; CTL2.  
 OS Synthetic.  
 PN MO9527507-A1.  
 PD 19-OCT-1995.  
 PE 06-APR-1995; U03989.  
 PR 06-APR-1994; US-223842.  
 PR 05-APR-1995; US-417210.  
 PA (VIR-) VIROGENETICS CORP.  
 PI Cox W; Paolletti E; Tartaglia J;  
 DR WPI; 95-366231/47.  
 DR N-PSDB; T04705.  
 PT Virulence-attenuated virus encoding an immunodeficiency virus  
 PT epitope - based on Copenhagen strain of vaccinia virus, used in the  
 PT prevention and treatment of diseases, e.g. vaccination against HIV  
 PT (Example 16; Fig 18; 208pp; English).  
 PS This sequence is a PC5POLT5A-derived CTL pool epitope. PC5POLT5A is  
 CC a plasmid contg. attenuated virus ALVAC recombinant expressing 3 CTL  
 CC pol epitopes, HIV1 gag (+pro) (IIIB) and gp120 (MN) and transmembrane  
 CC region. ALVAC-based recombinant viruses expressing extrinsic immunogens  
 CC as efficacious as vaccine vectors. Attenuated recombinant viruses such  
 CC as ALVAC or NYVAC can be engineered to comprise exogenous DNA in a non-  
 CC essential region of their genome. The exogenous DNA encodes at least one  
 CC immunodeficiency virus epitope. Such attenuated viruses (as above) and  
 CC derived antigens and antibodies are used in the prevention, therapy  
 CC and diagnosis of diseases. DNA from the recombinant viruses can be used  
 CC as probes or for generating primers or for immunisation. Attenuated,  
 CC recombinant viruses have enhanced safety making them safer for use in  
 CC vaccines.  
 SQ Sequence 49 AA;

Query Match 85.4%; Score 70; DB 16; Length 49;  
 Best Local Similarity 100.0%; Pred. No. 6.65e-01;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 16 PKVKQWPL 23  
 |||||  
 QY 2 PKVKQWPL 9

RESULT 7  
 ID R29709 standard; Protein; 522 AA.  
 AC R29709;  
 DT 26-APR-1993 (first entry)  
 DE p41 gag protein from htlr.  
 KW express proteins; lymphadenopathy syndrome; AIDS; HIV; HTLV;  
 KM serological immunoassays; antibodies to htlr; monoclonal antibodies;  
 OS Probes; ss.  
 KW Human T cell lymphotropic retrovirus.  
 KM EP-518443-A.  
 PN 16-DEC-1992.  
 PF 30-OCT-1985; 201711.

PR 31-OCT-1984; US-667501.  
 PR 30-JAN-1985; US-696534.  
 PR 06-SEP-1985; US-773447.  
 PA (CHIR) CHIRON CORP.

PI Barr PJ, Dina D, George-Nascimento C, Halliwell R;  
 PI Luciw PA, Parkes D, Pescador RS, Steimer K, Truett M;  
 DR WPI; 92-417329/51.  
 DR N-PSDB; Q31938.

PT Recombinant DNA construct including replication system recognised  
 PT by unicellular microorganism - used to form recombinant proteins  
 PT for diagnosing AIDS and lymphadenopathy syndrome  
 PS Example 11; Fig 5; 32pp; English.  
 CC This sequence was decoded from the p41 gag gene from htlr DNA.  
 CC Proteins associated with lymphadenopathy syndrome and/or AIDS may  
 CC be used in serological immunoassays to detect antibodies to htlr.  
 CC The polypeptides can be used alone or in fusion constructs to  
 CC produce antisera or monoclonal antibodies which may be used for  
 CC therapy or diagnosis.  
 SQ Sequence 522 AA;

Query Match 85.4%; Score 70; DB 6; Length 522;  
 Best Local Similarity 100.0%; Pred. No. 6.65e-01;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 174 PKVKQWPL 181  
 |||||  
 QY 2 PKVKQWPL 9

RESULT 8  
 ID P61510 standard; Protein; 543 AA.  
 AC P61510;  
 DT 13-SEP-1991 (first entry)  
 DE Sequence of pol protein encoded by ARV-2 cDNA cloned in pGAG41-10  
 DE for producing the fusion protein p41 gag.  
 KW LAV; HIV; ARV; HTLV; vaccine; AIDS; immunoassay; diagnosis;  
 KM lymphadenopathy syndrome.  
 OS Human T-cell lymphotropic virus III.  
 PN EP-181150-A.  
 PD 14-MAY-1986.  
 PF 30-OCT-1985; 307860.  
 PR 31-OCT-1984; US-667501.  
 PR 30-JAN-1985; US-696534.  
 PR 06-SEP-1985; US-773447.  
 PA (CHIR) CHIRON CORP.  
 PI Luciw PA, Dina D, Steimer K, Pescador RS, George-Nascimento C,  
 PI Parkes D, Halliwell R, Barr PJ, Truett M;  
 DR WPI; 86-126568/20.  
 DR N-PSDB; N60142.

PT New recombinant human T-cell lymphotropic retro virus proteins -  
 PT useful in diagnostic immunoassays for antibodies in humans, and  
 PT in prodn. of monoclonal antibodies, as vaccines etc.  
 PS Disclosure: Fig 5; 67pp; English.

CC The inventors claim a DNA construct contg. a DNA sequence  
 CC substantially as set forth in N60141, N60142, N60143, N60144, which  
 CC are each derived from AIDS-associated retroviruses. For the purposes  
 CC of this application, HTLV-III, LAV and ARV are generically referred  
 CC to as human T-cell lymphotropic retrovirus (htlr). The following  
 CC recombinant polypeptides are also claimed: (a) ARV-2 p16 gag;  
 CC (b) ARV-2 p25 gag; (c) ARV-2 env; (d) ARV-2 p31 pol. pGAG was  
 CC constructed from plasmid pGAG25-10 by inserting an SphI-HpaI  
 CC fragment from the ARV-2 genome containing the sequences from the  
 CC C-terminal p16 gag portion of the p53 gag precursor polypeptide and  
 CC part of the p25 gag protein between the SphI and BamHI sites of  
 CC pGAG25-10.  
 SQ Sequence 543 AA;

Query Match 85.4%; Score 70; DB 3; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 6.65e-01;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 174 PKVKQWPL 181  
 |||||

QY 2 PKVKQWPL 9

RESULT 9  
ID R08053 standard; protein: 912 AA.  
AC R08053;  
DT 18-JAN-1991 (first entry)  
DE ACNPV-HIVK-pol protein of HIVXB2 virus.  
KW HIV diagnosis; ACNPV-HIVK-pol; vaccine; HIVXB2;  
OS Human immunodeficiency virus - 1.  
PN WO9010230-A.  
PD 07-SEP-1990.  
PF 23-FEB-1990; CA0062.  
PR 18-APR-1989; GB-008725.  
PA (UYOT-) UNIV OF OTTAWA.  
PI Kang CY;  
DR N-PSDB: 005979.  
PT Improved polypeptide reagent for HIV diagnosis and vaccine -  
comprises portions of all 4 enzymes encoded by HIV-pol gene  
PS Disclosure: Page 11-23; 37pp; English.  
CC Recombinant ACNPV-HIVK-pol omits NH2-terminal sequences encoding the  
proteolytic active site of the HIV-pol protease. (Compare with ACNPV-  
HIVK-pol (006644) comprising the whole DNA sequence of the HIV-pol  
gene) When this sequence is expressed, the resulting gene product  
is not "processed", i.e. the 95 kD protein, comprising HIV-pol  
reverse transcriptase, HIV-pol RNase H and HIV-pol integrase,  
remains intact.  
CC An improved polypeptide reagent comprises portions of all of the  
4 enzymes, and is used in a diagnostic test for HIV infection.  
CC The peptide is also used in vaccines.  
SQ Sequence 912 AA;

Query Match 85.4%; Score 70; DB 2; Length 912;

Best Local Similarity 100.0%; Pred. No. 6,65e-01;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 83 PKVKQWPL 90  
|||||||

QY 2 PKVKQWPL 9

RESULT 10  
ID R08057 standard; protein: 912 AA.  
AC R08057;  
DT 18-JAN-1991 (first entry)  
DE HIV-1 pol protein of HIVMN.  
KW HIV diagnosis; HIV-pol; vaccine; HIVMN;  
OS Human immunodeficiency virus - 1.  
PN WO9010230-A.  
PD 07-SEP-1990.  
PF 23-FEB-1990; CA0062.  
PR 18-APR-1989; GB-008725.  
PA (UYOT-) UNIV OF OTTAWA.  
PI Kang CY;  
DR WPI: 90-290460/38.  
PT Improved polypeptide reagent for HIV diagnosis and vaccine -  
comprises portions of all 4 enzymes encoded by HIV-pol gene  
PS Disclosure: Page 11-23; 37pp; English.  
CC Several strains of HIV-1 were cloned and the corresponding amino  
acid sequence derived from the determined DNA sequences.  
CC An improved polypeptide reagent comprises portions of all of the  
4 enzymes, and is used in a diagnostic test for HIV infection.  
CC The peptide is also used in vaccines.  
SQ Sequence 912 AA;

Query Match 85.4%; Score 70; DB 2; Length 912;  
Best Local Similarity 100.0%; Pred. No. 6,65e-01;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;DB 83 PKVKQWPL 90  
|||||||

QY 2 PKVKQWPL 9

RESULT 11  
ID R09301 standard; protein: 982 AA.  
AC R09301;  
DT 27-FEB-1991 (first entry)  
DE Sequence deduced from pol gene of HIV 1-NDK.  
OS Human immunodeficiency virus; AIDS.  
PN WO9013630-A.  
PD 15-NOV-1990.  
PF 02-MAY-1990; F00312.  
PR 03-MAY-1989; FR-005914.  
PA (INRM) INSERM INST NAT SANTE.  
PI Barre-Sinoussi F, Chermann JC, Deaux C, Rey F, Sire J;  
PI Spire B; 90-361470/48.  
DR WPI: 90-361470/48.  
DR N-PSDB: Q06635.  
PT New HIV-NDK retrovirus and protein component - used in vaccines  
against immunodeficiency disorders and in raising Mabs for  
retro-virus detection in vivo.  
PS Disclosure: fig 2; 37pp; French.  
CC The HIV NDK virus was isolated from peripheral blood lymphocytes of  
an AIDS patient. A genomic library was prep'd. from DNA extracted  
from CEM cells infected with the virus. The library was screened  
with a pBT1 probe corresp. to a fragment from HIV 1. The virus is  
more cytopathic than other strains and is not inhibited by OKT4A.  
CC It has been deposited as CNCM I-857. The sequence can be used to  
express proteins useful for diagnosing the presence of NDK and  
related viruses and in vaccines against immunodeficiency diseases.  
CC See also R09301-5.  
SQ Sequence 982 AA;

Query Match 85.4%; Score 70; DB 2; Length 982;

Best Local Similarity 100.0%; Pred. No. 6,65e-01;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 153 PKVKQWPL 160  
|||||||

QY 2 PKVKQWPL 9

RESULT 12  
ID R12256 standard; protein: 1001 AA.  
AC R12256;  
DT 20-AUG-1991 (first entry)  
DE HIV-1 strain OYI POL protein.  
KW HIV-1; AIDS; retroviruses.  
OS Homo sapiens.  
PN US5019510-A.  
PD 28-MAY-1991.  
PR 28-OCT-1987; 113655.  
PR 28-OCT-1987; US-113655.  
PA (INSP) INST PASTEUR.  
PI Main-Hobson S, Huot T, Delaporte E, Brun-Vezinet F;  
DR WPI: 91-177518/24.  
PT Purified human retrovirus - is mutant of HIV-1 having  
characteristics of HIV-1 OYI, used in diagnosis of HIV infection  
PS Disclosure: fig 2; 23pp; English.  
CC This sequence constitutes the POL protein constituent of a new strain  
of HIV-1 retrovirus, OYI. This mutant retroviral strain is useful in  
an assay for diagnosing HIV infection. See also Q11943 (OYI  
nucleotide sequence), R12255 and R12257-62 (other HIV OYI constituent  
proteins).  
SQ Sequence 1001 AA;

Query Match 85.4%; Score 70; DB 3; Length 1001;  
Best Local Similarity 100.0%; Pred. No. 6,65e-01;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Db 173 pkvkwpl 180  
 |||||  
 QY 2 PKVKOMPL 9

RESULT 13  
 ID P61508 standard; Protein: 1003 AA.  
 AC P61508;  
 DT 19-AUG-1991 (first entry)  
 DE Sequence of ARV-2 (9B) pol protein.  
 KW LAV; HIV; ARV; HTLV; vaccine; AIDS; immunosassay; diagnosis;  
 OS Lymphadenopathy syndrome.  
 OS Human T-cell lymphotropic virus III.  
 PN EP-181150-A.  
 PD 14-MAY-1986.  
 PF 30-OCT-1985; 307860.  
 PR 31-OCT-1984; US-667501.  
 PR 30-JAN-1985; US-696534.  
 PR 06-SEP-1985; US-773447.  
 PA (CHIR) CHIRON CORP.  
 PI Luciw PA, Dina D, Steimer K, Pescador RS, George-Nascimento C,  
 PI Parkes D, Hallewell R, Barr PJ, Truett M;  
 DR N-PSDB; N60140.  
 PT New recombinant human T-cell lymphotropic retro virus proteins -  
 PT useful in diagnostic immunoassays for antibodies in humans, and  
 PT in prodn. of monoclonal antibodies, as vaccines etc.  
 PS Example; Fig 2; 67pp; English.  
 CC The inventors claim a DNA construct contg. a DNA sequence  
 CC substantially as set forth in N60141, N60142, N60143, N60144, which  
 CC are each derived from AIDS-associated retroviruses. For the purposes  
 CC of this application, HTLV-III, LAV and ARV are generically referred  
 CC to as human T-cell lymphotropic retrovirus (HTLV). The following  
 CC recombinant polypeptides are also claimed: (a) ARV-2 p16 gag;  
 CC (b) ARV-2 p25 gag; (c) ARV-2 env; (d) ARV-2 p31 pol.  
 SQ Sequence. 1003 AA;

Query Match 85.4%; Score 70; DB 3; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 6,65e-01;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 pkvkwpl 181  
 |||||  
 QY 2 PKVKOMPL 9

RESULT 14  
 ID R08061 standard; protein: 1003 AA.  
 AC R08061;  
 DT 18-JAN-1991 (first entry)  
 DE HIV-1 pol protein of HIVELI.  
 KW HIV diagnosis; HIV-pol; vaccine; HIVELI;  
 KW protein processing; reverse transcriptase; RNase; integrase.  
 OS Human immunodeficiency virus - 1.  
 PN WO9010230-A.  
 PD 07-SEP-1990.  
 PF 23-FEB-1990; CA0062.  
 PR 18-APR-1989; GB-008725.  
 PA (UYOT-) UNIV OF OTTAWA.  
 PI Kang CY;  
 DR WPI: 90-290460/38.  
 PT Improved polypeptide reagent for HIV diagnosis and vaccine -  
 PT comprises portions of all 4 enzymes encoded by HIV-pol gene  
 PS Disclosure; Page 11-23; 37pp; English.  
 CC Several strains of HIV-1 were cloned and the corresponding amino  
 CC acid sequence derived from the determined DNA sequences.  
 CC An improved polypeptide reagent comprises portions of all of the  
 CC 4 enzymes, and is used in a diagnostic test for HIV infection.  
 CC The peptide is also used in vaccines.  
 CC See also R08053-63.  
 SQ Sequence 1003 AA;

Query Match 85.4%; Score 70; DB 2; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 6,65e-01;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 pkvkwpl 181  
 |||||  
 QY 2 PKVKOMPL 9


RESULT 15  
 ID P60420 standard; Protein: 1003 AA.  
 AC P60420;  
 DT 20-AUG-1991 (first entry)  
 DE Sequence of LAV virus pol protein.  
 KW AIDS vaccine; diagnosis; immunosassay; HIV; HTLV-III.  
 OS Lymphadenopathy virus.  
 PN WO8602383-A.  
 PD 24-APR-1986.  
 PF 18-OCT-1985; E00548.  
 PR 18-OCT-1984; FR-016013.  
 PR 16-NOV-1984; GB-029099.  
 PR 21-JAN-1985; GB-001473.  
 PA (CNRS) CNRS CENT NAT RECH SCI.  
 PA (INSP) INRS PASTEUR.  
 PI Montagnier L, Krust B, Charnaret S, Clavel F, Chermann J-C,  
 PI Barre-Sinoussi F, Alizon M, Sonigo P, Stewart C, Dano O,  
 PI Wain-Hobson S;  
 DR WPI: 86-119166/18.  
 DR N-PSDB; N60365.  
 PT Purified glyco:protein and peptide(s) - are recognised by sera contg.  
 PT antibodies against lymphadenopathy virus and useful in detecting  
 PT AIDS antibodies or in vaccines.  
 PS Disclosure; Fig 4; 73pp; English.  
 CC The inventors claim a polypeptide which is recognised by sera of  
 CC human origin contg. antibodies against the virus of  
 CC lymphadenopathies (LAV) or acquired immune deficiency syndrome  
 CC (AIDS). Also claimed are various peptides corresp. to the AA  
 CC sequences deducible from proteins encoded by LAV DNA, defined by  
 CC specific residues (e.g. 12-32, 37-46, 49-79, 88-153) in accordance  
 CC with a formula given in the specification.  
 SQ Sequence 1003 AA;

Query Match 85.4%; Score 70; DB 3; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 6,65e-01;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 pkvkwpl 181  
 |||||  
 QY 2 PKVKOMPL 9

Search completed: Fri Sep 11 12:41:52 1998  
 Job time : 16 secs.


















(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

Mvsicon_pp      protein - protein database search, using Smith-Waterman algorithm
Run on:         Fri Sep 11 12:42:11 1998;      MasPar time 3.38 seconds
Tabular output not generated.                  97.203 Million cell updates/sec

```

```
Title: >US-08-452-843-2
Description: (1-9) from US08452843.pep
Perfect Score: 82
Sequence: 1 YPKVQWPL 9
```

Scoring table: PAM 150  
Gap 15

Searched: 120441 seqs, 36531193 residues

```
post-processing: Minimum Match 08
! Listing first 45 summaries
```

1:pir1 2:pir2 3:pir3 4:pir4 5:nr13d

Statistics: Mean 24.174; Variance 38.425; scale 0.629

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description           | Pred. No. |
|------------|-------|-------------|--------|----|--------|-----------------------|-----------|
| 1          | 70    | 85.4        | 61     | 5  | 1HWB1  | Hiv-1 reverse transcr | 1.95e-02  |
| 2          | 70    | 85.4        | 63     | 5  | 1HWAI  | Hiv-1 reverse transcr | 1.95e-02  |
| 3          | 70    | 85.4        | 83     | 5  | 1RTB1  | Hiv-1 reverse transcr | 1.95e-02  |
| 4          | 70    | 85.4        | 83     | 5  | 1RTB1  | Hiv-1 reverse transcr | 1.95e-02  |
| 5          | 70    | 85.4        | 84     | 5  | 1VRTB1 | Hiv-1 reverse transcr | 1.95e-02  |
| 6          | 70    | 85.4        | 84     | 5  | 1RTB1  | Hiv-1 reverse transcr | 1.95e-02  |
| 7          | 70    | 85.4        | 85     | 5  | 1RTB1  | Hiv-1 reverse transcr | 1.95e-02  |
| 8          | 70    | 85.4        | 151    | 2  | S67371 | pol polyprotein - hum | 1.95e-02  |
| 9          | 70    | 85.4        | 160    | 2  | S67333 | pol polyprotein - hum | 1.95e-02  |
| 10         | 70    | 85.4        | 214    | 5  | 1VRUB1 | Hiv-1 reverse transcr | 1.95e-02  |
| 11         | 70    | 85.4        | 214    | 5  | 1REUB1 | Hiv-1 reverse transcr | 1.95e-02  |
| 12         | 70    | 85.4        | 216    | 5  | 1RTB1  | Hiv-1 reverse transcr | 1.95e-02  |
| 13         | 70    | 85.4        | 218    | 5  | 1DIOB1 | Immunodeficiency vtru | 1.95e-02  |
| 14         | 70    | 85.4        | 219    | 2  | S33047 | RNA-directed DNA poly | 1.95e-02  |
| 15         | 70    | 85.4        | 219    | 2  | S33098 | RNA-directed DNA poly | 1.95e-02  |
| 16         | 70    | 85.4        | 219    | 2  | S33073 | RNA-directed DNA poly | 1.95e-02  |
| 17         | 70    | 85.4        | 219    | 2  | S33159 | RNA-directed DNA poly | 1.95e-02  |
| 18         | 70    | 85.4        | 219    | 2  | S33077 | RNA-directed DNA poly | 1.95e-02  |
| 19         | 70    | 85.4        | 219    | 2  | S33092 | RNA-directed DNA poly | 1.95e-02  |
| 20         | 70    | 85.4        | 219    | 2  | S33227 | RNA-directed DNA poly | 1.95e-02  |
| 21         | 70    | 85.4        | 219    | 2  | S33069 | RNA-directed DNA poly | 1.95e-02  |
| 22         | 70    | 85.4        | 219    | 2  | S32140 | RNA-directed DNA poly | 1.95e-02  |
| 23         | 70    | 85.4        | 224    | 5  | ZHYTBI | RNA-directed DNA poly | 1.95e-02  |

|    |    |      |     |   |        |                       |          |
|----|----|------|-----|---|--------|-----------------------|----------|
| 24 | 70 | 85.4 | 224 | 5 | 3HVTB1 | RNA-directed DNA poly | 1.95e-02 |
| 25 | 70 | 85.4 | 426 | 5 | IHMIB  | RNA-directed DNA poly | 1.95e-02 |
| 26 | 70 | 85.4 | 427 | 5 | IHNIB  | Immunodeficiency vtru | 1.95e-02 |
| 27 | 70 | 85.4 | 427 | 5 | IOWBB  | reverse transcriptase | 1.95e-02 |
| 28 | 70 | 85.4 | 427 | 5 | IHNVB  | hiv-1 reverse transcr | 1.95e-02 |
| 29 | 70 | 85.4 | 440 | 5 | IYVBL  | reverse transcriptase | 1.95e-02 |
| 30 | 70 | 85.4 | 427 | 5 | IYVTL  | hiv-1 reverse transcr | 1.95e-02 |
| 31 | 70 | 85.4 | 440 | 5 | IYVRL  | hiv-1 reverse transcr | 1.95e-02 |
| 32 | 70 | 85.4 | 441 | 5 | IYVRL  | hiv-1 reverse transcr | 1.95e-02 |
| 33 | 70 | 85.4 | 539 | 5 | IYVLA  | hiv-1 reverse transcr | 1.95e-02 |
| 34 | 70 | 85.4 | 543 | 5 | IYVLA  | hiv-1 reverse transcr | 1.95e-02 |
| 35 | 70 | 85.4 | 543 | 5 | IYVLA  | hiv-1 reverse transcr | 1.95e-02 |
| 36 | 70 | 85.4 | 543 | 5 | IYVLA  | hiv-1 reverse transcr | 1.95e-02 |
| 37 | 70 | 85.4 | 543 | 5 | IYVLA  | hiv-1 reverse transcr | 1.95e-02 |
| 38 | 70 | 85.4 | 554 | 5 | 2HYTA  | RNA-directed DNA poly | 1.95e-02 |
| 39 | 70 | 85.4 | 555 | 5 | 3HYTA  | RNA-directed DNA poly | 1.95e-02 |
| 40 | 70 | 85.4 | 556 | 5 | IDLOA  | Immunodeficiency vtru | 1.95e-02 |
| 41 | 70 | 85.4 | 556 | 5 | IHMIA  | RNA-directed DNA poly | 1.95e-02 |
| 42 | 70 | 85.4 | 558 | 5 | IHMIA  | Immunodeficiency vtru | 1.95e-02 |
| 43 | 70 | 85.4 | 558 | 5 | IHMVA  | hiv-1 reverse transcr | 1.95e-02 |
| 44 | 70 | 85.4 | 558 | 5 | IHMVA  | reverse transcriptase | 1.95e-02 |
| 45 | 70 | 85.4 | 558 | 5 | IYVRA  | reverse transcriptase | 1.95e-02 |

## ALIGNMENTS

|                       |    |          |  |  |
|-----------------------|----|----------|--|--|
| RESULT                | 1  | 1HMV1    | #type fragment   |  |
| ENTRY                 |    |          | Hiv-1 reverse transcriptase (EC 2.7.7.49), chain B, fragment |  |
| TITLE                 |    |          | 1 - human immunodeficiency virus type 1 (isolate BH10)       |  |
| ORGANISM              |    |          | #format_name human immunodeficiency virus type 1, HIV-1      |  |
| #note                 |    |          | isolate BH10 expressed in Escherichia coli                   |  |
| REFERENCE             |    |          | A67005   |  |
| #authors              |    |          | Rodgers, D.W.; Gamblin, S.J.; Harris, B.A.; Ray, S.; Culp,   |  |
|                       |    |          | J.S.; Hellmig, B.; Woolf, D.J.; Debouck, C.; Harrison, S.C.  |  |
|                       |    |          | submitted to the Brookhaven Protein Data Bank, December 1994 |  |
| #submission           |    |          | cross-references pdb:1HMV                                    |  |
| REFERENCE             |    |          | S05643   |  |
| #authors              |    |          | Mistahl, V.; Lazarus, G.M.; Miles, L.M.; Meyers, C.A.;       |  |
|                       |    |          | Debouck, C.  |  |
| #journal              |    |          | Arch. Biochem. Biophys. (1989) 273:347                       |  |
| #title                |    |          | Recombinant hiv-1 reverse transcriptase: purification,       |  |
|                       |    |          | primary structure, and polymerase/ribonuclease h             |  |
|                       |    |          | activities.  |  |
| #cross-references     |    |          | M01D:89372892  |  |
| COMMENT               |    |          | Resolution: 3.2 angstroms                                    |  |
| COMMENT               |    |          | Determination: X-ray diffraction                             |  |
| KEYWORDS              |    |          | Nucleotidyltransferase                                       |  |
| FEATURE               |    |          |  |  |
| 23-40                 |    |          | #region helix (right hand alpha)\                            |  |
| 42-45,53-61           |    |          | #region beta sheet   |  |
| SUMMARY               |    |          | #length 61 #checksum 5760                                    |  |
|                       |    |          |  |  |
| Query Match           |    |          | 85.4%; Score 70; DB 5; Length 61;                            |  |
| Best Local Similarity |    |          | 100.0%; Pred. No. 1,95e-02;                                  |  |
| Matches               |    |          | 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;           |  |
| Db                    | 15 | PKVKQMP1 | 22   |  |
|                       |    |          |  |  |
| QY                    | 2  | PKVKQMP1 | 9  |  |
| RESULT                | 2  | 1HMV1    | #type fragment   |  |
| ENTRY                 |    |          | Hiv-1 reverse transcriptase (EC 2.7.7.49), chain A, fragment |  |
| TITLE                 |    |          | 1 - human immunodeficiency virus type 1 (isolate BH10)       |  |
| ORGANISM              |    |          | #format_name human immunodeficiency virus type 1, HIV-1      |  |
| #note                 |    |          | isolate BH10 expressed in Escherichia coli                   |  |
| REFERENCE             |    |          | A67005   |  |
| #authors              |    |          | Rodgers, D.W.; Gamblin, S.J.; Harris, B.A.; Ray, S.; Culp,   |  |
|                       |    |          | J.S.; Hellmig, B.; Woolf, D.J.; Debouck, C.; Harrison, S.C.  |  |
|                       |    |          | submitted to the Brookhaven Protein Data Bank, December 1994 |  |
| #submission           |    |          | cross-references pdb:1HMV                                    |  |

REFERENCE S05643  
#authors Misrahi, V.; Lazarus, G.M.; Miles, L.M.; Meyers, C.A.;  
#journal Arch. Biochem. Biophys. (1989) 273:347  
#title Recombinant hiv-1 reverse transcriptase: purification,  
primary structure, and polymerase/ribonuclease h  
activities.  
#cross-references MUID:89372892  
COMMENT Resolution: 3.2 angstroms  
DETERMINATION: X-ray diffraction  
KEYWORDS Nucleotidyltransferase  
PDB-FILE 1.95e-02  
#region helix (right hand alpha)  
#length 63 #checksum 6742  
SUMMARY  
Query Match: 85.4%; Score 70; DB 5; Length 63;  
Best Local Similarity 100.0%; Pred. No. 1.95e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 19 PKVOMPL 26  
QY 2 PKVOMPL 9  
RESULT 3  
ENTRY 1RTBI #type fragment  
TITLE hiv-1 reverse transcriptase (EC 2.7.7.49), chain B, fragment  
1 - synthetic  
ALTERNATE\_NAMES  
ORGANISM #formal\_name synthetic  
#note strain hxb2 isolate, expressed in Escherichia coli  
A65560  
REFERENCE  
#authors Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby,  
I.; Keeling, J.; Dardy, G.; Jones, Y.; Stuart, D.;  
Stammers, D.  
#submission submitted to the Brookhaven Protein Data Bank, May 1995  
#cross-references PDB:1RTI  
REFERENCE  
#authors Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Kirby, C.R.I.;  
Keeling, J.; Dardy, G.; Jones, Y.; Stuart, D.I.; Stammers,  
D.  
#journal Nat. Struct. Biol. (1995) 2:293  
#title High resolution structures of hiv-1 rt from four rt-inhibitor  
complexes.  
REFERENCE  
#authors Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.;  
Stuart, D.  
#journal Nat. Struct. Biol. (1995) 2:303  
#title Mechanism of inhibition of hiv-1 reverse transcriptase by  
non-nucleoside inhibitors.  
REFERENCE  
#authors Stammers, D.K.; Somers, D.; Ross, C.K.; Kirby, I.; Ray, P.H.;  
Wilson, J.E.; Norman, M.; Ren, J.S.; Esnouf, R.M.; Garman,  
E.F.; Jones, E.Y.; Stuart, D.I.  
#journal J. Mol. Biol. (1994) 242:586  
#title Crystals of hiv-1 reverse transcriptase diffracting to 2.2  
angstrom resolution.  
COMMENT Resolution: 3.0 angstroms  
DETERMINATION: X-ray diffraction  
R-value: 0.236  
FEATURE  
KEYWORDS hiv-1 reverse transcriptase; nucleotidyltransferase  
PDB-FILE  
#region helix (right hand alpha)\n#region helix (right hand alpha)\n#region beta sheet\n#region beta sheet  
#length 83 #checksum 2890  
SUMMARY  
Query Match: 85.4%; Score 70; DB 5; Length 83;  
Best Local Similarity 100.0%; Pred. No. 1.95e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 15 PKVOMPL 22

QY 2 PKVOMPL 9  
RESULT 4  
ENTRY 1RTBI #type fragment  
TITLE hiv-1 reverse transcriptase (EC 2.7.7.49), chain B, fragment  
1 - hiv-1  
ALTERNATE\_NAMES  
PDB-FILE  
ORGANISM #formal\_name Human immunodeficiency virus #common\_name hiv-1  
#note strain hxb2 isolate, expressed in Escherichia coli  
A65555  
REFERENCE  
#authors Ren, J.; Esnouf, R.; Hopkins, A.; Willcox, B.; Jones, Y.;  
Ross, C.; Stammers, D.; Stuart, D.  
#submission submitted to the Brookhaven Protein Data Bank, March 1996  
#cross-references PDB:1RT1  
REFERENCE  
#authors Hopkins, A.L.; Ren, J.; Esnouf, R.M.; Willcox, B.E.; Jones,  
E.Y.; Ross, C.; Miyasaka, T.; Walker, R.T.; Tanaka, H.;  
Stammers, D.K.; Stuart, D.I.  
#journal J. Med. Chem. (1996) 39:1589  
#title Complexes of hiv-1 reverse transcriptase with inhibitors of  
the hept series reveal conformational changes relevant to  
the design of potent non-nucleoside inhibitors.  
REFERENCE  
#authors Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.;  
Stuart, D.  
#journal Nat. Struct. Biol. (1995) 2:303  
#title Mechanism of inhibition of hiv-1 reverse transcriptase by  
non-nucleoside inhibitors.  
REFERENCE  
#authors Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby,  
I.; Keeling, J.; Dardy, G.; Jones, Y.; Stuart, D.  
#journal Nat. Struct. Biol. (1995) 2:293  
#title High resolution structures of hiv-1 rt from four rt-inhibitor  
complexes.  
REFERENCE  
#authors Esnouf, R.; Hopkins, A.; Ross, C.; Jones, Y.;  
Stammers, D.; Stuart, D.  
#journal Structure (London) (1995) 3:915  
#title The structure of hiv-1 reverse transcriptase complexed with  
9-chloro-1-thio: lessons for inhibitor design.  
REFERENCE  
#authors Stammers, D.K.; Somers, D.O.; Ross, C.K.; Kirby, I.; Ray,  
P.H.; Wilson, J.E.; Norman, M.; Ren, J.S.; Esnouf, R.M.;  
Garman, E.F.  
#journal J. Mol. Biol. (1994) 242:586  
#title Crystals of hiv-1 reverse transcriptase diffracting to 2.2 a  
resolution.  
COMMENT Resolution: 2.55 angstroms  
DETERMINATION: X-ray diffraction  
R-value: 0.197  
FEATURE  
KEYWORDS hiv-1 reverse transcriptase; nucleotidyltransferase  
PDB-FILE  
#region helix (right hand alpha)\n#region helix (right hand alpha)\n#region beta sheet\n#region beta sheet  
#length 83 #checksum 2525  
SUMMARY  
Query Match: 85.4%; Score 70; DB 5; Length 83;  
Best Local Similarity 100.0%; Pred. No. 1.95e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 14 PKVOMPL 21  
QY 2 PKVOMPL 9  
RESULT 5  
ENTRY 1VRTBI #type fragment

```

TITLE      hiv-1 reverse transcriptase (EC 2.7.7.49), chain B, fragment
ALTERNATE_NAMES 1: synthetic
ORGANISM    #formal_name synthetic
            #note strain hxb2 isolate, expressed in Escherichia coli
            #A66891
REFERENCE   #authors
            Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby,
            I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.;
            Stammers, D.
#submission submitted to the Brookhaven Protein Data Bank, April 1995
#cross-references PDB:1VRT
REFERENCE   #authors
            Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Kirby, C.R.I.;
            Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.I.; Stammers,
            D.
#journal    Nat. Struct. Biol. (1995) 2:293
#title      High resolution structures of hiv-1 rt from four rt-inhibitor
            complexes.
REFERENCE   #authors
            Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.;
            Stuart, D.
#journal    Nat. Struct. Biol. (1995) 2:303
#title      Mechanism of inhibition of hiv-1 reverse transcriptase by
            non-nucleoside inhibitors.
REFERENCE   #authors
            Stammers, D.K.; Somers, D.; Ross, C.K.; Kirby, I.; Ray, P.H.;
            Wilson, J.E.; Norman, M.; Ren, J.S.; Esnouf, R.M.; Garman,
            E.F.; Jones, E.Y.; Stuart, D.I.
#journal    J. Mol. Biol. (1994) 242:586
#title      Crystals of hiv-1 reverse transcriptase diffracting to 2.2
            angstrom resolution.
COMMENT     Resolution: 2.2 angstroms
COMMENT     Determination: X-ray diffraction
COMMENT     R-value: 0.186
KEYWORDS    hiv-1 reverse transcriptase; nucleotidyltransferase
FEATURE     24-39-
            74-79-
            43-45-
            56-60,67-71-
SUMMARY     #length 84 #checksum 5239

Query Match      85.4%; Score 70; DB 5; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.95e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      15 PRVKOMPL 22
OY      2 PRVKOMPL 9

RESULT    6
ENTRY     1RT2B1 #type fragment
TITLE     hiv-1 reverse transcriptase (EC 2.7.7.49), chain B, fragment
            1 - hiv-1
ALTERNATE_NAMES
            hiv-1 rt
ORGANISM  #formal_name Human immunodeficiency virus #common name hiv-1
            #note strain hxb2 isolate, expressed in Escherichia coli
            A66556
REFERENCE #authors
            Ren, J.; Esnouf, R.; Hopkins, A.; Wilcox, B.; Jones, Y.;
            Ross, C.; Stammers, D.; Stuart, D.
#submission submitted to the Brookhaven Protein Data Bank, March 1996
#cross-references PDB:1RT2
REFERENCE   #authors
            Hopkins, A.L.; Ren, J.; Esnouf, R.M.; Wilcox, B.E.; Jones,
            E.Y.; Ross, C.; Miyasaka, T.; Walker, R.T.; Tanaka, H.;
            Stammers, D.K.; Stuart, D.I.
#journal    J. Med. Chem. (1996) 39:1589
#title      Complexes of hiv-1 reverse transcriptase with inhibitors of
            the hept serial reveal conformational changes relevant to
            the design of potent non-nucleoside inhibitors.

REFERENCE   #authors
            Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.;
            Stuart, D.
#journal    Nat. Struct. Biol. (1995) 2:303
#title      Mechanism of inhibition of hiv-1 reverse transcriptase by
            non-nucleoside inhibitors.
REFERENCE   #authors
            Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby,
            I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.
#journal    Nat. Struct. Biol. (1995) 2:293
#title      High resolution structures of hiv-1 rt from four rt-inhibitor
            complexes.
REFERENCE   #authors
            Ren, J.; Esnouf, R.; Hopkins, A.; Ross, C.; Jones, Y.;
            Stammers, D.; Stuart, D.
#journal    Structure (London) (1995) 3:915
#title      The structure of hiv-1 reverse transcriptase complexed with
            9-chloro-tlbo: lessons for inhibitor design.
REFERENCE   #authors
            Stammers, D.K.; Somers, D.O.; Ross, C.K.; Kirby, I.; Ray,
            P.H.; Wilson, J.E.; Norman, M.; Ren, J.S.; Esnouf, R.M.;
            Garman, E.F.
#journal    J. Mol. Biol. (1994) 242:586
#title      Crystals of hiv-1 reverse transcriptase diffracting to 2.2 a
            resolution.
COMMENT     Resolution: 2.55 angstroms
COMMENT     Determination: X-ray diffraction
COMMENT     R-value: 0.207
KEYWORDS    hiv-1 reverse transcriptase; nucleotidyltransferase
FEATURE     23-38-
            73-78-
            54-59,66-71-
SUMMARY     #length 84 #checksum 4388

Query Match      85.4%; Score 70; DB 5; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.95e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      14 PRVKOMPL 21
OY      2 PRVKOMPL 9

RESULT    7
ENTRY     1RT2B1 #type fragment
TITLE     hiv-1 reverse transcriptase (EC 2.7.7.49), chain B, fragment
            1 - synthetic
ALTERNATE_NAMES
            hiv-1 rt
ORGANISM  #formal_name synthetic
            #note strain hxb2 isolate, expressed in Escherichia coli
            A66559
REFERENCE #authors
            Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby,
            I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.;
            Stammers, D.
#submission submitted to the Brookhaven Protein Data Bank, May 1995
#cross-references PDB:1RT2
REFERENCE   #authors
            Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Kirby, C.R.I.;
            Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.I.; Stammers,
            D.
#journal    Nat. Struct. Biol. (1995) 2:293
#title      High resolution structures of hiv-1 rt from four rt-inhibitor
            complexes.
REFERENCE   #authors
            Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.;
            Stuart, D.
#journal    Nat. Struct. Biol. (1995) 2:303
#title      Mechanism of inhibition of hiv-1 reverse transcriptase by
            non-nucleoside inhibitors.
REFERENCE   #authors
            Stammers, D.K.; Somers, D.; Ross, C.K.; Kirby, I.; Ray, P.H.;
            Wilson, J.E.; Norman, M.; Ren, J.S.; Esnouf, R.M.; Garman,

```

#journal E.F. Jones, E.Y. Stuart, D.I.  
#title J. Mol. Biol. (1994) 242:586  
#comment Resolution: 2.2 angstroms  
#comment Determination: X-ray diffraction  
#comment R-value: 0.214  
#keywords hiv-1 reverse transcriptase; nucleotidyltransferase  
#feature  
26-41 #region helix (right hand alpha)\  
76-81 #region helix (right hand alpha)\  
#region beta sheet  
#summary  
58-61, 70-73 #length 85 #checksum 6905  
Query Match 85.4%; Score 70; DB 5; Length 85;  
Best Local Similarity 100.0%; Pred. No. 1.95e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 17 PKVKOMPL 24  
OY 2 PKVKOMPL 9  
RESULT 8  
ENTRY S63731 #type fragment  
TITLE pol polypeptide - human immunodeficiency virus type 1  
CONTAINS (isolate RJ9434) (fragment)  
retropepsin (EC 3.4.23.16); RNA-directed DNA polymerase (EC 2.7.7.49)  
ORGANISM #formal name human immunodeficiency virus type 1, HIV-1  
#variety isolate RJ9434  
DATE 20-Jul-1996 #sequence\_revision 27-Feb-1997 #text\_change 08-Sep-1997  
ACCESSIONS S63731  
REFERENCE S63731  
#authors Yamaguchi, K.  
#submission submitted to the EMBL Data Library, July 1995  
#accession S63731  
#molecule\_type DNA  
#residues 1-151 #label YAM  
#cross-references EMBL:U31385; NID:9961548; PID:9961549  
#experimental\_source isolate RJ9434  
REFERENCE S63703  
#authors Yamaguchi, K.; Byrn, R.A.  
#journal Biochim. Biophys. Acta (1995) 1253:136-140  
#title Clinical isolates of HIV-1 contain few pre-existing  
proteinase inhibitor resistance-conferring mutations.  
#accession S63703  
#status nucleic acid sequence not shown  
#molecule\_type DNA  
#residues 23-121 #label YAM  
#cross-references EMBL:U31385  
#experimental\_source isolate RJ9434  
GENETICS  
#gene pol  
#classification #superfamily pol polypeptide  
#keywords AIDS; aspartic proteinase; hydrolase; immunodeficiency;  
nucleotidyltransferase; polypeptide  
#feature  
23-121 #product:retropepsin #status predicted #label RTP  
#summary  
#length 151 #checksum 5994  
Query Match 85.4%; Score 70; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.95e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 140 PKVKOMPL 147  
OY 2 PKVKOMPL 9  
RESULT 9  
ENTRY S63753 #type fragment

TITLE pol polypeptide - human immunodeficiency virus type 1  
CONTAINS (isolate RJ9533M) (fragment)  
retropepsin (EC 3.4.23.16); RNA-directed DNA polymerase (EC 2.7.7.49)  
ORGANISM #formal name human immunodeficiency virus type 1, HIV-1  
#variety isolate RJ9533M  
DATE 20-Jul-1996 #sequence\_revision 27-Feb-1997 #text\_change 08-Sep-1997  
ACCESSIONS S63753  
REFERENCE S63753  
#authors Yamaguchi, K.  
#submission submitted to the EMBL Data Library, July 1995  
#accession S63753  
#molecule\_type DNA  
#residues 1-160 #label YAM  
#cross-references EMBL:U31409; NID:9961596; PID:9961597  
#experimental\_source isolate RJ9533M  
REFERENCE S63703  
#authors Yamaguchi, K.; Byrn, R.A.  
#journal Biochim. Biophys. Acta (1995) 1253:136-140  
#title Clinical isolates of HIV-1 contain few pre-existing  
proteinase inhibitor resistance-conferring mutations.  
#accession S63727  
#status nucleic acid sequence not shown  
#molecule\_type DNA  
#residues 32-130 #label YAM  
#cross-references EMBL:U31409  
#experimental\_source isolate RJ9533M  
GENETICS  
#gene pol  
#classification #superfamily pol polypeptide  
#keywords AIDS; aspartic proteinase; hydrolase; immunodeficiency;  
nucleotidyltransferase; polypeptide  
#feature  
32-130 #product:retropepsin #status predicted #label RTP  
#summary  
#length 160 #checksum 5613  
Query Match 85.4%; Score 70; DB 2; Length 160;  
Best Local Similarity 100.0%; Pred. No. 1.95e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 149 PKVKOMPL 156  
OY 2 PKVKOMPL 9  
RESULT 10  
ENTRY 1VRUB1 #type fragment  
TITLE hiv-1 reverse transcriptase (EC 2.7.7.49), chain B, fragment  
1 synthetic  
ALTERNATE\_NAMES hiv-1 rt  
ORGANISM #formal name synthetic  
#note strain hxb2 isolate, expressed in Escherichia coli  
REFERENCE A66892  
#authors Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Dardy, G.; Jones, Y.; Stuart, D.;  
Stammers, D.  
#submission submitted to the Brookhaven Protein Data Bank, April 1995  
#cross-references PDB:1VRU  
REFERENCE TND32393  
#authors Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Kirby, C.R.I.;  
Keeling, J.; Dardy, G.; Jones, Y.; Stuart, D.I.; Stammers, D.  
#journal Nat. Struct. Biol. (1995) 2:293  
#title High resolution structures of hiv-1 rt from four rt-inhibitor complexes.  
REFERENCE TND32394  
#authors Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.;  
Stuart, D.  
#journal Nat. Struct. Biol. (1995) 2:303  
#title Mechanism of inhibition of hiv-1 reverse transcriptase by non-nucleoside inhibitors.  
REFERENCE TND32395

#authors                   Stammers, D.K.; Somers, D.; Ross, C.K.; Kirby, I.; Ray, P.H.;  
                             Wilson, J.E.; Norman, M.; Ren, J.S.; Esnouf, R.M.; Garman,  
                             E.F.; Jones, E.Y.; Stuart, D.I.  
#journal                   J. Mol. Biol. (1994) 242:586  
#title                     Crystals of hiv-1 reverse transcriptase diffracting to 2.2  
                             angstrom resolution.  
COMMENT                   Resolution: 2.4 angstroms  
COMMENT                   Determination: X-ray diffraction  
KEYWORDS                  R-value: 0.187  
                             hiv-1 reverse transcriptase; nucleotidyltransferase  
FEATURE  
#   25-40                   #region helix (right hand alpha)\  
75-80                   #region helix (right hand alpha)\  
95-99                   #region helix (right hand 3-10)\  
109-114                  #region helix (right hand 3-10)\  
122-125                  #region helix (right hand 3-10)\  
132-134                  #region helix (right hand 3-10)\  
151-164                  #region helix (right hand alpha)\  
166-171                  #region helix (right hand alpha)\  
192-210                  #region helix (right hand alpha)\  
44-46,139-143,  
127-129                  #region beta sheet\  
57-61, 68-72            #region beta sheet\  
102-107,183-188,  
176-180                  #region beta sheet\  
SUMMARY                  #length 214 #checks 4548  
                             #region beta sheet  
Query Match              85.4%; Score 70; DB 5; Length 214;  
Best Local Similarity 100.0%; Pred. No. 1,95e-02;  
Matches                  8; Conservative   0; Mismatches   0; Indels   0; Gaps   0;  
                             2 PRVKOMPL 9

Db   16 PRVKOMPL 23  
QY   2 PRVKOMPL 9

RESULT 11  
ENTRY                   1REV1           #type fragment  
TITLE                   hiv-1 reverse transcriptase (EC 2.7.7.49), chain B, fragment  
                             1 - Human immunodeficiency virus type 1  
ALTERNATE\_NAMES        hiv-1 rt  
PDB TITLE              hiv-1 reverse transcriptase  
ORGANISM               #formal\_name Human immunodeficiency virus type 1  
                             strain hxb2 isolate, expressed in Escherichia coli  
REFERENCE              A66504  
#authors               Ren, J.; Esnouf, R.; Hopkins, A.; Ross, C.; Jones, Y.;  
                             Stammers, D.; Stuart, D.  
#submission            submitted to the Brookhaven Protein Data Bank, September 1995  
REFERENCE              #cross-references PDB:1REV  
                             TNO34372  
#journal                Ren, J.; Esnouf, R.; Hopkins, A.; Ross, C.; Jones, Y.;  
                             Stammers, D.; Stuart, D.  
#title                  Structure (London) (1995) 3:915  
                             The structure of hiv-1 reverse transcriptase complexed with  
                             9-chloro-1lbo: lessons for inhibitor design.  
#journal                TNO34373  
#authors                Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.;  
                             Stuart, D.  
REFERENCE              Nat. Struct. Biol. (1995) 2:303  
                             Mechanism of inhibition of hiv-1 reverse transcriptase by  
                             non-nucleoside inhibitors.  
#journal                TNO34374  
#authors                Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby,  
                             I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.  
#title                  Nat. Struct. Biol. (1995) 2:293  
                             High resolution structures of hiv-1 rt from four rt-inhibitor  
                             complexes.  
REFERENCE              TNO34375  
#journal                Stammers, D.K.; Somers, D.O.; Ross, C.K.; Kirby, I.; Ray, P.H.;  
                             Garman, E.F.  
#authors                P.H.; Wilson, J.E.; Norman, M.; Ren, J.S.; Esnouf, R.M.;  
                             J. Mol. Biol. (1994) 242:586  
#title                  Crystals of hiv-1 reverse transcriptase diffracting to 2.2 a

COMMENT                   Resolution.  
COMMENT                   Resolution: 2.6 angstroms  
COMMENT                   Determination: X-ray diffraction  
KEYWORDS                  R-value: 0.224  
                             aids; aspartyl protease endonuclease; hydrolase;  
                             nucleotidyltransferase; nucleotidyltransferase hiv-1  
                             reverse transcriptase; polypeptide  
FEATURE  
#   27-42                   #region helix (right hand alpha)\  
77-82                   #region helix (right hand alpha)\  
97-101                   #region helix (right hand 3-10)\  
110-116                  #region helix (right hand 3-10)\  
123-127                  #region helix (right hand 3-10)\  
134-136                  #region helix (right hand 3-10)\  
153-173                  #region helix (right hand 3-10)\  
194-211                  #region helix (right hand alpha)\  
46-48,141-145,  
129-131                  #region beta sheet\  
58-63,70-75            #region beta sheet\  
104-109,185-190,  
177-182                  #region beta sheet\  
SUMMARY                  #length 214 #checks 3320  
                             #region beta sheet  
Query Match              85.4%; Score 70; DB 5; Length 214;  
Best Local Similarity 100.0%; Pred. No. 1,95e-02;  
Matches                  8; Conservative   0; Mismatches   0; Indels   0; Gaps   0;  
                             2 PRVKOMPL 9

Db   18 PRVKOMPL 25  
QY   2 PRVKOMPL 9

RESULT 12  
ENTRY                   1RTV1           #type fragment  
TITLE                   hiv-1 reverse transcriptase (EC 2.7.7.49), chain B, fragment  
                             1 - synthetic  
ALTERNATE\_NAMES        hiv-1 rt  
PDB TITLE              hiv-1 reverse transcriptase  
ORGANISM               #formal\_name synthetic  
                             strain hxb2 isolate, expressed in Escherichia coli  
REFERENCE              A66561  
#authors               Ren, J.; Esnouf, R.; Ross, C.; Jones, Y.; Stammers, D.;  
                             Stuart, D.  
#submission            submitted to the Brookhaven Protein Data Bank, May 1995  
REFERENCE              #cross-references PDB:1RTV  
                             TNO30518  
#journal                Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.;  
                             Stuart, D.  
#title                  Nat. Struct. Biol. (1995) 2:303  
                             Mechanism of inhibition of hiv-1 reverse transcriptase by  
                             non-nucleoside inhibitors.  
#journal                TNO30519  
#authors                Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Kirby, C.R.I.;  
                             Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.I.; Stammers,  
                             D.  
REFERENCE              Nat. Struct. Biol. (1995) 2:293  
                             High resolution structures of hiv-1 rt from four rt-inhibitor  
                             complexes.  
#journal                TNO30520  
#authors                Stammers, D.K.; Somers, D.; Ross, C.K.; Kirby, I.; Ray, P.H.;  
                             Wilson, J.E.; Norman, M.; Ren, J.S.; Esnouf, R.M.; Garman,  
                             E.F.; Jones, E.Y.; Stuart, D.I.  
#title                  J. Mol. Biol. (1994) 242:586  
                             Crystals of hiv-1 reverse transcriptase diffracting to 2.2  
                             angstrom resolution.  
COMMENT                   Resolution: 2.35 angstroms  
COMMENT                   Determination: X-ray diffraction  
KEYWORDS                  R-value: 0.219  
                             hiv-1 reverse transcriptase; nucleotidyltransferase  
FEATURE  
#   27-42                   #region helix (right hand alpha)\  
77-82                   #region helix (right hand alpha)\  
111-114                  #region helix (right hand 3-10)\  
124-127                  #region helix (right hand 3-10)\

|                       |             |                     |        |               |
|-----------------------|-------------|---------------------|--------|---------------|
| Query Match           | 85.43;      | Score 70;           | DB 2;  | Length 219;   |
| Best Local Similarity | 100.08;     | Pred. NO. 1.85e-02; |        |               |
| Matches               | 8;          | Conservative        | 0;     | Mismatches 0; |
|                       |             |                     | Indels | 0;            |
|                       |             |                     | Gaps   | 0;            |
| Db                    | 19 PKVKOMPL | 26                  |        |               |
|                       |             |                     |        |               |
|                       |             |                     |        |               |
|                       | 2 PKVKOMPL  | 9                   |        |               |
|                       |             |                     |        |               |
|                       |             |                     |        |               |



SUN SEP 13 10:56:18 1998

US-08-452-843-2.rpr

Page 7

Search completed: Fri Sep 11 12:42:41 1998  
Job time : 30 secs

**THIS PAGE BLANK (USPTO)**

Release 3.1A John F. Collins, Bloccomputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

Mpsrch_pp    protein - protein database search, using smith-Waterman algorithm

Run on:      Fri Sep 11 12:42:58 1998;      MspPar time 2.38 seconds
Tabular output not generated.              94,986 Million cell updates/sec

```

Title: >US-08-452-843-2  
Description: (1-9) from US08452843.pep  
Perfect Score: 82  
Sequence: 1 YPKVKQWPL 9

Scoring table: PAM 150  
Gap 15

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swiss1

Statistics: Mean 25.202; Variance 33.401; scale 0.7555

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result | No. | Score | Query | Length | DB | ID        | Description            | Pred.    | No. |
|--------|-----|-------|-------|--------|----|-----------|------------------------|----------|-----|
|        | 1   | 70    | 85.4  | 1002   | 1  | POL_HV1RH | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 2   | 70    | 85.4  | 1002   | 1  | POL_HV1EL | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 3   | 70    | 85.4  | 1002   | 1  | POL_HV1Z2 | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 4   | 70    | 85.4  | 1002   | 1  | POL_HV1U4 | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 5   | 70    | 85.4  | 1002   | 1  | POL_HV1ND | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 6   | 70    | 85.4  | 1003   | 1  | POL_HV1NS | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 7   | 70    | 85.4  | 1003   | 1  | POL_HV1A2 | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 8   | 70    | 85.4  | 1003   | 1  | POL_HV1Y2 | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 9   | 70    | 85.4  | 1003   | 1  | POL_HV1H2 | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 10  | 70    | 85.4  | 1003   | 1  | POL_HV1O1 | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 11  | 70    | 85.4  | 1006   | 1  | POL_HV1NM | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 12  | 70    | 85.4  | 1007   | 1  | POL_HV1JR | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 13  | 70    | 85.4  | 1015   | 1  | POL_HV1BR | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 14  | 70    | 85.4  | 1015   | 1  | POL_HV1VP | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 15  | 70    | 85.4  | 1015   | 1  | POL_HV1B5 | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 16  | 70    | 85.4  | 1015   | 1  | POL_HV1B1 | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 17  | 70    | 85.4  | 1027   | 1  | POL_S1VC7 | POL POLYPEPTIDE (PROTE | 1.94e-03 |     |
|        | 18  | 67    | 81.7  | 1027   | 1  | POL_HV1MA | POL POLYPEPTIDE (PROTE | 8.65e-03 |     |
|        | 19  | 67    | 81.7  | 1035   | 1  | POL_HV1ZR | POL POLYPEPTIDE (PROTE | 8.65e-03 |     |
|        | 20  | 65    | 79.3  | 1009   | 1  | POL_S1VGR | POL POLYPEPTIDE (PROTE | 2.30e-02 |     |
|        | 21  | 65    | 79.3  | 1058   | 1  | POL_HV1D2 | POL POLYPEPTIDE (PROTE | 2.30e-02 |     |
|        | 22  | 64    | 78.0  | 147    | 1  | YK42_MTCU | HYPOTHETICAL 17.0 KD P | 3.72e-02 |     |
|        | 23  | 64    | 78.0  | 1054   | 1  | POL_S1VMK | POL POLYPEPTIDE (PROTE | 3.72e-02 |     |

|    |    |      |      |   |           |                         |          |
|----|----|------|------|---|-----------|-------------------------|----------|
| 24 | 64 | 78.0 | 1056 | 1 | POL_STVW1 | POL. POLYURENEIN (PROTE | 3.73e-02 |
| 25 | 64 | 78.0 | 1057 | 1 | POL_STVW1 | POL. POLYURENEIN (PROTE | 6.03e-02 |
| 26 | 63 | 76.8 | 1124 | 1 | POL_STVW2 | POL. POLYPROTEIN (PROTE | 6.73e-02 |
| 27 | 61 | 74.4 | 305  | 1 | RNH_BP1   | RIBONUCLEASE H (EC 3.1  | 1.55e-01 |
| 28 | 61 | 74.4 | 1019 | 1 | POL_STV4  | POL. POLYURENEIN (PROTE | 1.55e-01 |
| 29 | 61 | 74.4 | 1022 | 1 | POL_STVAP | POL. POLYURENEIN (PROTE | 1.55e-01 |
| 30 | 61 | 74.4 | 1035 | 1 | POL_HV2AP | POL. POLYURENEIN (PROTE | 1.55e-01 |
| 31 | 61 | 74.4 | 1036 | 1 | POL_HV2B0 | POL. POLYURENEIN (PROTE | 1.55e-01 |
| 32 | 61 | 74.4 | 1047 | 1 | POL_STVW1 | POL. POLYPROTEIN (PROTE | 1.55e-01 |
| 33 | 61 | 74.4 | 1055 | 1 | POL_HV2ST | POL. POLYURENEIN (PROTE | 1.55e-01 |
| 34 | 61 | 74.4 | 1073 | 1 | POL_HV2D1 | POL. POLYURENEIN (PROTE | 1.55e-01 |
| 35 | 61 | 74.4 | 1124 | 1 | POL_STV5E | POL. POLYURENEIN (PROTE | 1.55e-01 |
| 36 | 61 | 74.4 | 1124 | 1 | POL_STV5E | POL. POLYURENEIN (PROTE | 1.55e-01 |
| 37 | 61 | 74.4 | 1142 | 1 | POL_STV2E | POL. POLYURENEIN (PROTE | 1.55e-01 |
| 38 | 60 | 73.2 | 1056 | 1 | POL_STV7  | POL. POLYPROTEIN (PROTE | 2.48e-01 |
| 39 | 58 | 70.7 | 1056 | 1 | POL_BIV06 | POL. POLYURENEIN (PROTE | 2.48e-01 |
| 40 | 58 | 70.7 | 229  | 1 | GTH_M9B4T | GLUTATHIONE S-TRANSFER  | 6.21e-01 |
| 41 | 58 | 70.7 | 1034 | 1 | POL_HV2A1 | POL. POLYURENEIN (PROTE | 6.21e-01 |
| 42 | 58 | 70.7 | 1049 | 1 | POL_HV2A1 | POL. POLYURENEIN (PROTE | 6.21e-01 |
| 43 | 58 | 70.7 | 1145 | 1 | POL_STVAV | POL. POLYURENEIN (PROTE | 6.21e-01 |
| 44 | 58 | 70.7 | 1145 | 1 | POL_STVAV | POL. POLYURENEIN (PROTE | 6.21e-01 |
| 45 | 58 | 70.7 | 1146 | 1 | POL_STVAC | POL. POLYPROTEIN (PROTE | 6.21e-01 |

## ALIGNMENTS

|                       |   |           |              |                    |
|-----------------------|---|-----------|--------------|--------------------|
| RESULT                | 1   | STANDARD: | PRT:         | 1002 AA.           |
| ID                    | POL_H1HRH   |           |              |                    |
| AC                    | P05959:   |           |              |                    |
| DT                    | 01-NOV-1988 (REL. 09, CREATED)                                      |           |              |                    |
| DT                    | 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)                         |           |              |                    |
| DT                    | 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)                       |           |              |                    |
| DE                    | POL POLYPROTEIN (PROTEASE (RETROPEPSIN) (EC 3.4.23.16): REVERSE     |           |              |                    |
| DE                    | TRANSCRIPTASE (EC 2.7.7.49): RIBONUCLEASE H (EC 3.1.26.4)).         |           |              |                    |
| GN                    | POL.  |           |              |                    |
| OS                    | HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (R/HA1 ISOLATE) (HIV-1).        |           |              |                    |
| OC                    | VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;   |           |              |                    |
| CC                    | LENTIVIRINAE.   |           |              |                    |
| CC                    | [1]   |           |              |                    |
| RP                    | SEQUENCE FROM N.A.  |           |              |                    |
| RA                    | STARBUCH B.R., HAHN B.H., SHAW G.M., MCNEELY P.D., MODROW S.,       |           |              |                    |
| RA                    | WOLF H.C., PARKS E.S., PARKS W.P., JOSEPHS S.F., GALLO R.C.,        |           |              |                    |
| RA                    | WONG-STALL F.;  |           |              |                    |
| RL                    | SUBMITTED (XXX-1987) TO THE HIV DATA BANK.                          |           |              |                    |
| CC                    | -1- PTM: CLEAVAGE SITES THAT YIELD THE NATURE PROTEINS REMAIN TO BE |           |              |                    |
| CC                    | DETERMINED.   |           |              |                    |
| CC                    | -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO   |           |              |                    |
| CC                    | KNOWN AS THE RETROPEPSIN FAMILY.                                    |           |              |                    |
| DR                    | EMBL; M17451; G328569; ..   |           |              |                    |
| DR                    | HSSP; P03366; 1HR.  |           |              |                    |
| DR                    | HIV; M17451; POLSRR.  |           |              |                    |
| DR                    | PROSITE; PS00141; ASP_PROTEASE; 1.                                  |           |              |                    |
| KW                    | AIDS; POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;      |           |              |                    |
| KW                    | RNA-DIRECTED DNA POLYMERASE.  |           |              |                    |
| FT                    | CHAIN   | 56        | 154          | PROTEASE.          |
| FT                    | ACT_SITE  | 80        | 80           | BY SIMILARITY.     |
| SO                    | SEQUENCE  | 1002 AA;  | 113755 MW;   | GDE2B1B2 CRC32;    |
| Query Match           |   |           |              |                    |
| Best Local Similarity |   | 85.4%;    | Score 70;    | DB 1; Length 1002; |
| Matches               |   | 8;        | Conservative | 0;                 |
| Mismatches            |   | 0;        | Indels       | 0;                 |
| Gaps                  |   | 0;        |              |                    |
| Db                    | 173 PKYQWPL 180   |           |              |                    |
|                       |   |           |              |                    |
| OY                    | 2 PKYQWPL 9   |           |              |                    |
| RESULT 2.             |   |           |              |                    |
| ID                    | POL_H1EYL   | STANDARD: | PRT:         | 1002 AA.           |
| AC                    | P04589:   |           |              |                    |
| DT                    | 13-AUG-1987 (REL. 05, CREATED)                                      |           |              |                    |
| DT                    | 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)                         |           |              |                    |
| DT                    | 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)                       |           |              |                    |



FT ACT SITE 80 BY SIMILARITY.  
 SO SEQUENCE 1002 AA: 113621 MM: FC05DF15F CRC32:  
 Query Match 85.4%; Score 70; DB 1; Length 1002;  
 Best Local Similarity 100.0%; Pred. No. 1.94e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 173 PKYKMWPL 180  
 |||||  
 QY 2 PKYKMWPL 9

RESULT 6 STANDARD: PRT: 1003 AA.

AC P12497,  
 DT 01-OCT-1989 (REL. 12, CREATED)  
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE POL POLYPEPTIDE (PROTEASE (RETROPEPSIN) (EC 3.4.23.16); REVERSE  
 DE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).  
 GN POL.  
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (NEW YORK-5 ISOLATE) (HIV-1).  
 OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 OC LENTIVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A. (CLONE PNL4-3).  
 RA BUCKNER C.E., BUCKNER-WHITE A.J., WILLEY R.L., MCCOY J.;  
 RA SUBMITTED (JUN-1988) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RL [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 57-155.  
 RX MEDLINE: 90354401.  
 RA FITZGERALD P.M.D., MCKEEVER B.M., VAN MIDDLESMORTH J.F.,  
 RA SPRINGER J.P., HELMBACH J.C., LEU C.-T., HERBER W.K., DIXON R.A.F.,  
 RA DARKE P.L.;  
 RL J. BIOL. CHEM. 265:14209-14219(1990).  
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 DR EMBL: M19921; G328419; -  
 DR PDB: SHVP; 15-OCT-91.  
 DR PDB: 4PHV; 31-OCT-93.  
 DR HIV: M19921; POLSND43.  
 DR PROSITE: PS00141; ASP\_PROTEASE: 1.  
 DR AIDS: POLYPEPTIDE: ASPARTYL PROTEASE; ENDONUCLEASE;  
 KW RNA-DIRECTED DNA POLYMERASE; 3D-STRUCTURE.  
 KM RNA-DIRECTED DNA POLYMERASE; 3D-STRUCTURE.  
 FT ACT\_SITE 81 81  
 FT STRAND 58 59  
 FT STRAND 66 71  
 FT TURN 72 73  
 FT STRAND 74 80  
 FT TURN 82 83  
 FT STRAND 88 90  
 FT STRAND 99 105  
 FT TURN 106 107  
 FT STRAND 108 122  
 FT TURN 123 124  
 FT STRAND 125 134  
 FT STRAND 140 141  
 FT HELIX 143 149  
 FT TURN 150 150  
 FT STRAND 152 154  
 SO SEQUENCE 1003 AA: 113535 MM: SED59879 CRC32:

Query Match 85.4%; Score 70; DB 1; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 1.94e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 174 PKYKMWPL 181  
 |||||  
 QY 2 PKYKMWPL 9

RESULT 7 STANDARD: PRT: 1003 AA.

ID POL\_HVIA2  
 AC P03369;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE POL POLYPEPTIDE (PROTEASE (RETROPEPSIN) (EC 3.4.23.16); REVERSE  
 DE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).  
 GN POL.  
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (ARV2/SF2 ISOLATE) (HIV-1).  
 OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 OC LENTIVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 85090453.  
 RA SANCHEZ-PESCADOR R., POWER M.D., BARR P.J., STEIMER K.S.,  
 RA STEMPEN M.M., BROWN-SHIMER S.L., GEE W.W., RENNARD A., RANDOLPH A.,  
 RA LEVY J.A., DINA D., LUCIF P.A.;  
 RL SCIENCE 227:484-492(1985).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 57-155.  
 RX MEDLINE: 89346747.  
 RA WLODAR A., MILLER M., JASKOLSKI M., SATHANARAYANA B.K.,  
 RA BALDWIN E., WEBER I.T., SELK L.M., CLAWSON L., SCHNEIDER J.,  
 RA KENT S.B.H.;  
 RL SCIENCE 245:616-621(1989).  
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 DR EMBL: K02007; G328662; -  
 DR PIR: A03968; GNVMA2.  
 DR PDB: 3HVP; 15-JAN-90.  
 DR HIV: K02007; POLSSE2.  
 DR PROSITE: PS00141; ASP\_PROTEASE: 1.  
 DR AIDS: POLYPEPTIDE: ASPARTYL PROTEASE; ENDONUCLEASE;  
 KW RNA-DIRECTED DNA POLYMERASE; 3D-STRUCTURE.  
 KM RNA-DIRECTED DNA POLYMERASE; 3D-STRUCTURE.  
 FT ACT\_SITE 81 81  
 FT STRAND 66 71  
 FT TURN 72 73  
 FT STRAND 74 80  
 FT STRAND 87 90  
 FT STRAND 98 103  
 FT TURN 106 107  
 FT STRAND 110 115  
 FT STRAND 119 122  
 FT TURN 123 124  
 FT STRAND 125 128  
 FT STRAND 131 134  
 FT STRAND 140 141  
 FT HELIX 143 149  
 FT TURN 150 150  
 SO SEQUENCE 1003 AA: 113723 MM: A94EB76C CRC32:

Query Match 85.4%; Score 70; DB 1; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 1.94e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 174 PKYKMWPL 181  
 |||||  
 QY 2 PKYKMWPL 9

RESULT 8 STANDARD: PRT: 1003 AA.

ID POL\_HVIA2  
 AC P35963;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE POL POLYPEPTIDE (PROTEASE (RETROPEPSIN) (EC 3.4.23.16); REVERSE  
 DE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).

GN POL. HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (YU-2 ISOLATE) (HIV-1).  
OS VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
OC LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 93021387.  
RA LI Y., HUI H., BURGESS C.J., PRICE R.W., SHARP P.M., HAHN B.H.,  
RA SHAW G.M.;  
RL J. VIROL. 66:6587-6600(1992).  
CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
DR EMBL: M93258; -; NOT\_ANNOTATED\_CDS.  
DR PIR: B44001; B44001.  
DR HSSP: P03366; 1HRH.  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
RW AIDS: POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;  
RW RNA-DIRECTED DNA POLYMERASE.  
FT CHAIN 57 155  
FT ACT\_SITE 82 82 BY SIMILARITY.  
SQ SEQUENCE 1003 AA; 113794 MW; 99272DE9 CRC32;  
  
Query Match 85.4%; Score 70; DB 1; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 1.94e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
DB 174 PRVKOMPL 181  
OY 2 PRVKOMPL 9  
RESULT 9  
ID POL.HV1H2 STANDARD; PRT: 1003 AA.  
AC P04585;  
DT 13-AUG-1987 (REL. 09, CREATED)  
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE POLYPROTEIN (PROTEASE (RETROPEPSIN) (EC 3.4.23.16); REVERSE  
DE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).  
GN POL.  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HXB2 ISOLATE) (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
OC LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87299196.  
RA RATNER L., FISHER A., JAGODZINSKI L.L., MITSUYA H., LIU R.-S.,  
RA GALLO R.C., WONG-STAL F.;  
RL AIDS RES. HUM. RETROVIRUSES 3:57-69(1987).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 156-595.  
RX MEDLINE: 96097398.  
RA REN J., ESNOUF R., HOPKINS A., ROSS C., JONES Y., STAMMERS D.,  
RA STUART D.;  
RL STRUCTURE 3:915-926(1995).  
CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
DR EMBL: K03455; G327746; ALT\_INIT.  
DR PDB: 1REV; 14-OCT-96.  
DR HIV: K03455; POLSHX82.  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
RW AIDS: POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;  
RW RNA-DIRECTED DNA POLYMERASE; 3D-STRUCTURE.  
FT CHAIN 57 155  
FT ACT\_SITE 81 81 BY SIMILARITY.  
SQ SEQUENCE 1003 AA; 113769 MW; 4AFD4B80 CRC32;  
  
Query Match 85.4%; Score 70; DB 1; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 1.94e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
DB 174 PRVKOMPL 181  
OY 2 PRVKOMPL 9  
RESULT 10  
ID POL.HV1OY STANDARD; PRT: 1003 AA.  
AC P20892;  
DT 01-FEB-1991 (REL. 17, CREATED)  
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE POLYPROTEIN (PROTEASE (RETROPEPSIN) (EC 3.4.23.16); REVERSE  
DE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).  
GN POL.  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (OY1 ISOLATE) (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
OC LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 9018544.  
RA HUET T., DAZZA M.C., BRUN-VEZINET F., ROELANTS G.E., MAIN-HOBSON S.,  
RL AIDS 3:707-715(1989).  
CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- THE OY1 ISOLATE WAS TAKEN FROM THE BLOOD OF A HEALTHY GABONESE  
CC INDIVIDUAL.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
DR EMBL: M26727; G328444; -.  
DR HSSP: P03366; 1HRH.  
DR HIV: M26727; POLSOY1.  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
RW AIDS: POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;  
RW RNA-DIRECTED DNA POLYMERASE.  
FT CHAIN 57 155  
FT ACT\_SITE 81 81 BY SIMILARITY.  
SQ SEQUENCE 1003 AA; 113718 MW; E50B705E CRC32;  
  
Query Match 85.4%; Score 70; DB 1; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 1.94e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
DB 174 PRVKOMPL 181  
OY 2 PRVKOMPL 9  
RESULT 11  
ID POL.HV1MN STANDARD; PRT: 1006 AA.  
AC P05961;  
DT 01-NOV-1988 (REL. 09, CREATED)  
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE POLYPROTEIN (PROTEASE (RETROPEPSIN) (EC 3.4.23.16); REVERSE  
DE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).  
GN POL.  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (MN ISOLATE) (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
OC LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 88219542.  
RA GURGO C., GUD H.-G., FRANCHINI G., ALDOVINI A., COLALATI E.,  
RA FARRELL K., WONG-STAL F., GALLO R.C., REITZ M.S. JR.;  
RL VIROLOGY 164:531-536(1989).  
CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- THE MN ISOLATE WAS TAKEN FROM A PEDIATRIC AIDS PATIENT IN 1984.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
DR EMBL: M17449; -; NOT\_ANNOTATED\_CDS.

DR HSP; P03366; 1HRH.  
 DR HIV; M17449; POLSMN.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 KW AIDS; POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;  
 KW RNA-DIRECTED DNA POLYMERASE.  
 FT CHAIN 60 158  
 FT ACT\_SITE 84 84  
 FT SITE 565 565 BY SIMILARITY.  
 FT SITE 565 565 IN-FRAME TERMINATION CODON.  
 SQ SEQUENCE 1006 AA; 113860 MW; 70477EC0 CRC32;

Query Match 85.4%; Score 70; DB 1; Length 1006;  
 Best Local Similarity 100.0%; Pred. No. 1.94e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 177 PKVKOMPL 184  
 |||||  
 2 PKVKOMPL 9

RESULT 12  
 ID POL\_HV1JR STANDARD; PRT: 1007 AA.  
 AC P20875;  
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE POL POLYPROTEIN (PROTEASE (RETROPEPSIN) (EC 3.4.23.16); REVERSE  
 DE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).  
 GN POL.  
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (JRCSF ISOLATE) (HIV-1).  
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 OC LENTIVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA KOYANAGI S., CHEN I.S.Y.;  
 RA SUBMITTED (DEC-1988) TO THE HIV DATA BANK.  
 RL -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 CC EMBL; M38429; G327814; -.  
 DR HSP; P03367; 1AAO.  
 DR HIV; M38429; POLSJRCSE.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 KW AIDS; POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;  
 KW RNA-DIRECTED DNA POLYMERASE.  
 FT CHAIN 61 159  
 FT ACT\_SITE 85 85 BY SIMILARITY.  
 FT SITE 85 85  
 SQ SEQUENCE 1007 AA; 114081 MW; 492C03ED CRC32;

Query Match 85.4%; Score 70; DB 1; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 1.94e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 178 PKVKOMPL 185  
 |||||  
 2 PKVKOMPL 9

RESULT 13  
 ID POL\_HV1BR STANDARD; PRT: 1015 AA.  
 AC P03367;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE POL POLYPROTEIN (PROTEASE (RETROPEPSIN) (EC 3.4.23.16); REVERSE  
 DE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).  
 GN POL.  
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (BRU ISOLATE) (HIV-1).  
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 OC LENTIVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85099333.

RA MAIN-HOBSON S., SONIGO P., DANOS O., COLE S., ALIZON M.;  
 RL CELL 40:9-17(1985).  
 RN [2]  
 RP REVISIONS TO 23-35.  
 RA ALIZON M., MAIN-HOBSON S., MONTAGNIER L., SONIGO P.;  
 RL CELL 46:63-74(1986).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
 RX MEDLINE; 92190341.

RA SPINELLI S., LIU Q.Z., ALZARI P.M., HIREL P.H., POLJAK R.J.;  
 RL BIOCHIMIE 73:1391-1396(1991).  
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 CC EMBL; K02013; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A03966; GNMVLY.  
 DR PDB; 1HHP; 15-OCT-92.

DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 KW AIDS; POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;  
 KW RNA-DIRECTED DNA POLYMERASE; 3D-STRUCTURE.  
 FT CHAIN 69 167  
 FT ACT\_SITE 93 93 BY SIMILARITY.  
 FT STRAND 78 82  
 FT STRAND 87 92  
 FT TURN 94 95

FT STRAND 100 102  
 FT STRAND 111 117  
 FT TURN 118 119  
 FT STRAND 120 134  
 FT TURN 135 136  
 FT STRAND 137 146  
 FT STRAND 152 153  
 FT HELIX 155 161  
 FT TURN 162 162  
 SQ SEQUENCE 1015 AA; 115031 MW; F34C547E CRC32;

Query Match 85.4%; Score 70; DB 1; Length 1015;  
 Best Local Similarity 100.0%; Pred. No. 1.94e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 186 PKVKOMPL 193  
 |||||  
 2 PKVKOMPL 9

RESULT 14  
 ID POL\_HV1PV STANDARD; PRT: 1015 AA.  
 AC P03368;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE POL POLYPROTEIN (PROTEASE (RETROPEPSIN) (EC 3.4.23.16); REVERSE  
 DE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).  
 GN POL.  
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (PV22 ISOLATE) (HIV-1).  
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 OC LENTIVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 8511157.  
 RA MEDSING M.A., SMITH D.H., CABRADILLA C.D., BENTON C.V., LASKY L.A.,  
 RA CAPON D.J.;  
 RL NATURE 313:450-458(1985).  
 RN [2]  
 RP REVISION.  
 RA MEDSING M.A.;  
 RA SUBMITTED (XXX-1987) TO THE HIV DATA BANK.  
 RL -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.

DR EMBL; K02083; G328555; NOT\_ANNOTATED\_CDS.  
 DR PIR; A03967; GNMVWL.  
 DR HSSP; P03367; 1HHP.  
 DR HIV; K02083; POLSPV22.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 KM AIDS; POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;  
 KM RNA-DIRECTED DNA POLYMERASE.  
 FT CHAIN 69 167  
 FT ACT SITE 93 93 BY SIMILARITY  
 SQ SEQUENCE 1015 AA; 115090 MW; 28A1F7C8 CRC32;  
 Query Match 85.4%; Score 70; DB 1; Length 1015;  
 Best Local Similarity 100.0%; Pred. No. 1.94e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 186 PKYKMWPL 193  
 QY 2 PKYKMWPL 9  
 RESULT 15  
 ID POL\_HY1B1 STANDARD; PRT: 1015 AA.  
 AC P03366;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE POL POLYPROTEIN (PROTEASE (RETROPEPSIN) (EC 3.4.23.16); REVERSE  
 DE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).  
 GN POL.  
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (BH10 ISOLATE) (HIV-1).  
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 OC LENTIVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 8511123.  
 RA RATTNER L., HASELTINE W., PATARCA R., LIVAK K.J., STARCICH B.R.,  
 RA JOSEPHS S.F., DORAN E.R., RAFALSKI J.A., WHITEHORN E.A.,  
 RA BAUMEISTER K., IVANOFF L., PETTEWAY S.R. JR., PEARSON M.L.,  
 RA LAUTENBERGER J.A., PAPAS T.S., GHARAB J., CHANG N.T., GALLO R.C.,  
 RA WONG-STALL F.,  
 RA NATURE 313:277-284(1985).  
 RL [2]  
 RP 3D-STRUCTURE MODELLING OF PROTEASE DOMAIN.  
 RX MEDLINE; 89146134.  
 RA WEBER I.T., MILLER M., JASKOLSKI M., LEIS J., SKALKA A.M.,  
 RA WLODAWER A.,  
 RA SCIENCE 243:928-931(1989).  
 RL [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 69-167.  
 RX MEDLINE; 90044107.  
 RA LAPARTO R., BLINDELL T., HEMMINGS A., OVERINGTON J., WILDESPIN A.,  
 RA WOOD S., MERSON J.R., WHITTLE P.J., DANLEY D.E., GEOGHEGAN K.F.,  
 RA HAWRYLIK S.J., LEE S.E., SCHELD K.G., HOBART P.M.,  
 RA NATURE 342:299-302(1989).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 69-167.  
 RX MEDLINE; 90341771.  
 RA ERICSSON J., NEIDHART D.J., VANDRIE J., KEMPF D.J., WANG X.C.,  
 RA NORRICK D.W., PLATTNER J.J., RITTENHOUSE J.W., TURON M., WIDEBURG N.,  
 RA KOHLRENNER W.E., SIMMER R., HELFRICH R., PAUL D.A., KNIGGE M.,  
 RA SCIENCE 249:527-533(1990).  
 RL [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 594-729.  
 RX MEDLINE; 91188281.  
 RA DAVIES J.F. II, HOSTOMSKA Z., HOSTOMSKY Z., JORDAN S.R.,  
 RA MATTHEWS D.A.,  
 RA SCIENCE 252:88-95(1991).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 168-723.  
 RX MEDLINE; 93317673.  
 RA JACOBO-MOLINA A., DING J., NANNI R.G., CLARK A.D. JR., LU X.,  
 RA TANTILLO C., WILLIAMS R.L., KAMER G., FERRIS A.L., CLARK P., HIZI A.,

RA HUGHES S.H., ARNOLD E.,  
 RA PROC. NATL. ACAD. SCI. U.S.A. 90:6320-6324(1993).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 168-723.  
 RX MEDLINE; 92311654.  
 RA KOHLSTADT L.A., WANG J., FRIEDMAN J.M., RICE P.A., STEITZ T.A.,  
 RA SCIENCE 256:1783-1790(1992).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 168-723.  
 RA HSIOU Y., DING J., DAS K., CLARK A.D. JR., HUGHES S.H., ARNOLD E.,  
 RA SUBMITTED (APR-1996) TO THE PDB DATA BANK.  
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 CC EMBL; M15654; G326385; -.  
 DR PIR; A03965; GNMVH3.  
 DR PDB; 1HVP; 15-APR-92.  
 DR PDB; 9HVP; 15-JUL-92.  
 DR PDB; 1HRH; 15-OCT-94.  
 DR PDB; 3PHV; 15-JAN-92.  
 DR PDB; 1HMT; 31-JAN-94.  
 DR PDB; 1HOS; 31-OCT-93.  
 DR PDB; 3HVT; 15-OCT-94.  
 DR PDB; 1HV1; 30-APR-94.  
 DR PDB; 1HVJ; 30-APR-94.  
 DR PDB; 1HVK; 30-APR-94.  
 DR PDB; 1HVL; 30-APR-94.  
 DR PDB; 1HEP; 31-MAY-94.  
 DR PDB; 1HEG; 31-MAY-94.  
 DR PDB; 1HPS; 31-AUG-94.  
 DR PDB; 1HTE; 31-JUL-94.  
 DR PDB; 1HTE; 31-JUL-94.  
 DR PDB; 1HTG; 31-JUL-94.  
 DR PDB; 1SBG; 13-OCT-94.  
 DR PDB; 1DLO; 01-AUG-96.  
 DR PDB; 1GNN; 08-NOV-96.  
 DR PDB; 1GNO; 08-NOV-96.  
 DR HIV; M15654; POLSRH102.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 KW AIDS; POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;  
 KW RNA-DIRECTED DNA POLYMERASE; 3D-STRUCTURE.  
 RN CHAIN 69 167  
 FT ACT SITE 93 93 BY SIMILARITY.  
 FT CHAIN 167  
 FT ACT SITE 93 93  
 FT STRAND 70 71  
 FT STRAND 78 83  
 FT TURN 84 85  
 FT STRAND 86 92  
 FT TURN 94 95  
 FT STRAND 100 101  
 FT TURN 111 117  
 FT STRAND 118 119  
 FT STRAND 120 134  
 FT TURN 135 136  
 FT STRAND 137 145  
 FT STRAND 152 153  
 FT STRAND 155 158  
 FT HELIX 159 162  
 FT TURN 164 166  
 SQ SEQUENCE 1015 AA; 115021 MW; 26F6A003 CRC32;  
 Query Match 85.4%; Score 70; DB 1; Length 1015;  
 Best Local Similarity 100.0%; Pred. No. 1.94e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 186 PKYKMWPL 193  
 QY 2 PKYKMWPL 9  
 Search completed: Fri Sep 11 12:43:04 1998  
 Job time : 6 secs.



(TM)

Distribution rights by Oxford Molecular Ltd

protein - protein database search, using smith-waterman algorithm

03 380 Wi114 217 234  
MasPar Time 4.06 seconds

ues.

sptremb16

1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mhc 6:sp\_organelle 7:sp\_phage 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp\_vertebrate  
13:sp\_unclassified

62; scale 0.736

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

|        |   |              |                     |                 |               |
|--------|---|--------------|---------------------|-----------------|---------------|
| SO     | SEQUENCE  | 245 AA;      | 28738 MW;           | 48907759 CRC32; |               |
|        | Query Match   |              | 100.0%;             | Score 82;       | DB 11;        |
|        | Best Local Similarity   | 100.0%;      | Pred. No. 5.23e-06; |                 |               |
|        | Matches   | 9;           | Conservative        | 0;              | Mismatches 0; |
|        |   |              |                     |                 | Indels 0;     |
|        |   |              |                     |                 | Gaps 0        |
| Db     | 9 YPKVKOMPL 17  |              |                     |                 |               |
|        |   |              |                     |                 |               |
|        | 1 YPKVKOMPL 9   |              |                     |                 |               |
| QY     |   |              |                     |                 |               |
| RESULT | 2   |              |                     |                 |               |
| ID     | Q75832  | PRELIMINARY; | PRT;                | 245 AA.         |               |
| AC     | Q75832;   |              |                     |                 |               |
| DT     | 01-NOV-1996 (TREMBLREL. 01, CREATED)                              |              |                     |                 |               |
| DT     | 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)                 |              |                     |                 |               |
| DT     | 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)               |              |                     |                 |               |
| DE     | REVERSE TRANSCRIPTASE (FRAGMENT).                                 |              |                     |                 |               |
| GN     | HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).                      |              |                     |                 |               |
| OS     | VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE; |              |                     |                 |               |
| OC     | LENTIVIRINAE.   |              |                     |                 |               |
| ON     | [1]   |              |                     |                 |               |

RP SEQUENCE FROM N.A.  
RC STRAIN-ZH106 FROM AUSTRALIA;  
RA ZHENG N.N., HURREN L., NEILAN B.A., COOPER D.A., DELANEY S.F.,  
RA MOUJEN P.W.;  
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U64189; G1470284;  
KW RNA-DIRECTED DNA POLYMERASE.  
FT NON\_TER 1  
FT CHAIN 1  
SQ SEQUENCE 245 AA; 28584 MW; 16A34C26 CRC32;

1 Query Match 9 91.5%; Score 75; DB 11; Length 245;  
Best Local Similarity 88.9%; Pred. No. 2.08e-04;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 10 PKVKOMPL 18  
OY 1 PKVKOMPL 9

RESULT 3  
ID 072878 PRELIMINARY; PRT; 136 AA.  
AC 072878;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE POL POLYPROTEIN (FRAGMENT).  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
LN LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-2-3-107;  
RA YAMAGUCHI K., BYRN R.A.;  
RL BIOCHIM. BIOPHYS. ACTA 1253:136-140(1995).  
DR EMBL: U31402; G961583;  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
KW POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE.  
FT NON\_TER 1  
FT CHAIN 1  
FT NON\_TER 8 >106 PROTEASE.  
SQ SEQUENCE 136 AA; 14787 MW; 35B093BC CRC32;

Query Match 85.4%; Score 70; DB 11; Length 136;  
Best Local Similarity 100.0%; Pred. No. 2.66e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 125 PKVKOMPL 132  
OY 2 PKVKOMPL 9

RESULT 4  
ID 072868 PRELIMINARY; PRT; 140 AA.  
AC 072868;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE POL POLYPROTEIN (FRAGMENT).  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
LN LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RJ9622;  
RA YAMAGUCHI K., BYRN R.A.;  
RL BIOCHIM. BIOPHYS. ACTA 1253:136-140(1995).  
DR EMBL: U31392; G961563;  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
KW POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE.  
FT NON\_TER 1  
FT CHAIN 1  
FT NON\_TER 15 >113 PROTEASE.  
SQ SEQUENCE 140 AA; 140 MW; 140 PROTEASE.

SQ SEQUENCE 140 AA; 15182 MW; 0F9FD551 CRC32;

Query Match 85.4%; Score 70; DB 11; Length 140;  
Best Local Similarity 100.0%; Pred. No. 2.66e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 132 PKVKOMPL 139  
OY 2 PKVKOMPL 9

RESULT 5  
ID 072865 PRELIMINARY; PRT; 143 AA.  
AC 072865;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE POL POLYPROTEIN (FRAGMENT).  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
LN LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RJ9560;  
RA YAMAGUCHI K., BYRN R.A.;  
RL BIOCHIM. BIOPHYS. ACTA 1253:136-140(1995).  
DR EMBL: U31388; G961557;  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
KW POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE.  
FT NON\_TER 1  
FT CHAIN 1  
FT NON\_TER 15 >113 PROTEASE.  
SQ SEQUENCE 143 AA; 15683 MW; 84E91476 CRC32;

Query Match 85.4%; Score 70; DB 11; Length 143;  
Best Local Similarity 100.0%; Pred. No. 2.66e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 132 PKVKOMPL 139  
OY 2 PKVKOMPL 9

RESULT 6  
ID 072877 PRELIMINARY; PRT; 145 AA.  
AC 072877;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE POL POLYPROTEIN (FRAGMENT).  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
LN LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RJ14671;  
RA YAMAGUCHI K., BYRN R.A.;  
RL BIOCHIM. BIOPHYS. ACTA 1253:136-140(1995).  
DR EMBL: U31401; G961581;  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
KW POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE.  
FT NON\_TER 1  
FT CHAIN 1  
FT NON\_TER 19 >117 PROTEASE.  
SQ SEQUENCE 145 AA; 15671 MW; 1ACF95BE CRC32;

Query Match 85.4%; Score 70; DB 11; Length 145;  
Best Local Similarity 100.0%; Pred. No. 2.66e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 136 PKVKOMPL 143  
OY 2 PKVKOMPL 9

```

RESULT 7
ID 072864 PRELIMINARY: PRT: 149 AA.
AC 072864:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE POL. POLYPROTEIN (FRAGMENT).
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC LENTIVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RJ9532:
RA YAMAGUCHI K., BYRN R.A.:
RL BIOCHIM. BIOPHYS. ACTA 1253:136-140(1995).
DR EMBL: U31388; G961555;
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE.
FT NON_TER 1
FT CHAIN 21 >119
FT NON_TER 149 149
SQ SEQUENCE 149 AA; 16127 MW; 06771F46 CRC32;

Query Match
Best Local Similarity 100.0%; Score 70; DB 11; Length 149;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 138 PRVKOMPL 145
OY 2 PRVKOMPL 9

RESULT 8
ID 072885 PRELIMINARY: PRT: 160 AA.
AC 072885:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE POL. POLYPROTEIN (FRAGMENT).
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC LENTIVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RJ9533M;
RA YAMAGUCHI K., BYRN R.A.:
RL BIOCHIM. BIOPHYS. ACTA 1253:136-140(1995).
DR EMBL: U31409; G961597;
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE.
FT NON_TER 1
FT CHAIN 32 >130
FT NON_TER 160 160
SQ SEQUENCE 160 AA; 17491 MW; E17A111E CRC32;

Query Match
Best Local Similarity 100.0%; Score 70; DB 11; Length 160;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 149 PRVKOMPL 156
OY 2 PRVKOMPL 9

RESULT 9
ID 040468 PRELIMINARY: PRT: 259 AA.
AC 040468:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
DE HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).

```

```

OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC LENTIVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-L7;
RA SCHMIT J.C.:
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-L7;
RA SCHMIT J.C., COGNIAUX J., HERMANS P., VAN VAECK C., SPRECHER S.,
RA VAN REMOORTELT B., WITVROUW M., BALZARINI J., DESMYTER J., DE CLERCQ E.,
RA VANDAMME A.M.:
RL J. INFECT. DIS. 174:962-968(1996).
DR EMBL: AJ002376; E1168920;
KW RNA-DIRECTED DNA POLYMERASE.
FT NON_TER 1
FT NON_TER 259 259
SQ SEQUENCE 259 AA; 30023 MW; 6A48B7D0 CRC32;

Query Match
Best Local Similarity 100.0%; Score 70; DB 11; Length 259;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 19 PRVKOMPL 26
OY 2 PRVKOMPL 9

RESULT 10
ID 040467 PRELIMINARY: PRT: 259 AA.
AC 040467:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC LENTIVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-L6;
RA SCHMIT J.C.:
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-L6;
RA SCHMIT J.C., COGNIAUX J., HERMANS P., VAN VAECK C., SPRECHER S.,
RA VAN REMOORTELT B., WITVROUW M., BALZARINI J., DESMYTER J., DE CLERCQ E.,
RA VANDAMME A.M.:
RL J. INFECT. DIS. 174:962-968(1996).
DR EMBL: AJ002375; E1168918;
KW RNA-DIRECTED DNA POLYMERASE.
FT NON_TER 1
FT NON_TER 259 259
SQ SEQUENCE 259 AA; 29974 MW; 4048E1CF CRC32;

Query Match
Best Local Similarity 100.0%; Score 70; DB 11; Length 259;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 19 PRVKOMPL 26
OY 2 PRVKOMPL 9

RESULT 11
ID 040465 PRELIMINARY: PRT: 259 AA.
AC 040465:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE REVERSE TRANSCRIPTASE (FRAGMENT).

```

OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
OC LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-L4;  
RA SCHMITT J.C.;  
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-L4;  
RA SCHMITT J.C.;  
RL J. INJECT. DIS. 174:962-968(1996).  
DR EMBL: A1002373; E1168914; -;  
KM RNA-DIRECTED DNA POLYMERASE.  
FT NON\_TER 1 1  
FT CHAIN 100 >341  
SQ SEQUENCE 259 AA; 30111 MW; 125868A3 CRC32;  
Query Match 85.4%; Score 70; DB 11; Length 259;  
Best Local Similarity 100.0%; Pred. No. 2.66e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 19 PKVKOMPL 26  
OY 2 PKVKOMPL 9  
RESULT 12  
ID 039212 PRELIMINARY; PRT; 341 AA.  
AC 039212;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE POL. POLYPOLYMERASE (FRAGMENT).  
GN POL.  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
OC LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT A;  
RA WONG J.K., HEZAREH M., GUNTARD H.F., HAVLIR D.V., IGNACIO C.C.,  
RA SPINA C.A., RICHMAN D.D.;  
RL SCIENCE 278:1291-1295(1997).  
DR EMBL: AF027720; G2619082; -;  
KW PROSITE: PS00141; ASP\_PROTEASE; 1.  
KM POLYPOLYMERASE; HYDROLASE; ASPARTYL PROTEASE.  
FT NON\_TER 1 1  
FT CHAIN 100 >341  
FT CHAIN 341 341  
SQ SEQUENCE 341 AA; 38987 MW; D04A96D6 CRC32;  
Query Match 85.4%; Score 70; DB 11; Length 341;  
Best Local Similarity 100.0%; Pred. No. 2.66e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 118 PKVKOMPL 125  
OY 2 PKVKOMPL 9  
RESULT 13  
ID 039213 PRELIMINARY; PRT; 341 AA.  
AC 039213;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE POL. POLYPOLYMERASE (FRAGMENT).  
GN POL.

OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
OC LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT A;  
RA WONG J.K., HEZAREH M., GUNTARD H.F., HAVLIR D.V., IGNACIO C.C.,  
RA SPINA C.A., RICHMAN D.D.;  
RL SCIENCE 278:1291-1295(1997).  
DR EMBL: AF027721; G2619084; -;  
KW PROSITE: PS00141; ASP\_PROTEASE; 1.  
KM POLYPOLYMERASE; HYDROLASE; ASPARTYL PROTEASE.  
FT NON\_TER 1 1  
FT CHAIN 100 >341  
FT CHAIN 341 341  
SQ SEQUENCE 341 AA; 39016 MW; 207E0F04 CRC32;  
Query Match 85.4%; Score 70; DB 11; Length 341;  
Best Local Similarity 100.0%; Pred. No. 2.66e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 118 PKVKOMPL 125  
OY 2 PKVKOMPL 9  
RESULT 14  
ID 039200 PRELIMINARY; PRT; 341 AA.  
AC 039200;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE POL. POLYPOLYMERASE (FRAGMENT).  
GN POL.  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
OC LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT A;  
RA WONG J.K., HEZAREH M., GUNTARD H.F., HAVLIR D.V., IGNACIO C.C.,  
RA SPINA C.A., RICHMAN D.D.;  
RL SCIENCE 278:1291-1295(1997).  
DR EMBL: AF027708; G2619058; -;  
KW PROSITE: PS00141; ASP\_PROTEASE; 1.  
KM POLYPOLYMERASE; HYDROLASE; ASPARTYL PROTEASE.  
FT NON\_TER 1 1  
FT CHAIN 100 >341  
FT CHAIN 341 341  
SQ SEQUENCE 341 AA; 38942 MW; 1553D571 CRC32;  
Query Match 85.4%; Score 70; DB 11; Length 341;  
Best Local Similarity 100.0%; Pred. No. 2.66e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 118 PKVKOMPL 125  
OY 2 PKVKOMPL 9  
RESULT 15  
ID 039210 PRELIMINARY; PRT; 341 AA.  
AC 039210;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE POL. POLYPOLYMERASE (FRAGMENT).  
GN POL.  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
OC LENTIVIRINAE.

```
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN-PATIENT A;
RA      WONG J.K., HEZAREH M., GUNTARD H.F., HAVIR D.V., IGNACIO C.C.,
RA      SPINA C.A., RICHMAN D.D.;
RL      SCIENCE 278:1291-1295(1997).
DR      EMBL; AF027718; G2619078;
DR      PROSITE; PS00141; ASP_PROTEASE; 1.
KW      POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE.
FT      NON_TER      1      1
FT      CHAIN      1      >99      PROTEASE.
FT      CHAIN      100      >341      REVERSE TRANSCRIPTASE.
FT      NON_TER      341      341
SQ      SEQUENCE      341 AA; 38962 MM; AIDFOE7A CRC32;
```

Query Match 85.4%; Score 70; DB 11; Length 341;  
Best Local Similarity 100.0%; Pred. No. 2.66e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 118 PKWKOMPL 125  
QY 2 PKWKOMPL 9

Search completed: Fri Sep 11 12:43:43 1998  
Job time : 20 secs.

**THIS PAGE BLANK (USPTO)**

(W.T.)

Distribution rights by Oxford Molecular Ltd

1. **FRUKYAAAF** 9

## Gap 15

Existing first 45 summaries

29:part29

ved by analysis of the total score distribu-

4339 2044 61.1

9 AA;

9; Conse

РЕКЛАМА 9

RESULT 2  
ID W20553 standard; Protein: 326 AA.  
AC W20553;  
DT 04-AUG-1997 (first entry)  
DE H. pylori cell envelope inner membrane protein 6093906.aa.  
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacterium; life cycle; activator;  
KW bactericidal; inhibitor; duodenal ulcer disease; chronic gastritis;  
KW diagnosis.  
OS Helicobacter pylori.  
FH Key Location/Qualifiers  
FT misc\_difference 6 /label= unknown  
FT /note= "encoded by ARC"  
PN W09640893-A1.  
PD 19-DEC-1996.  
PE 06-JUN-1996; U09122.  
PR 07-JUN-1995; US-487032.  
PR 01-APR-1996; US-630405.  
PA (ASTR ) ASTRA AB.  
PI Berglindh OT, Smith D, Mellgaerd BL;  
DR WPI: 97-052306/05.  
DR N-PSDB: T67824.  
PT Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PS Claim 56; Page 707-708; 1481pp; English.  
CC This sequence shows a Helicobacter pylori cell envelope protein  
CC that may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
CC overlapping contigs generated by mechanically shearing the bacterial  
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
SQ Sequence 326 AA;

Query Match 69.4%; Score 50; DB 22; Length 326;  
Best Local Similarity 55.6%; Pred. No. 1.77e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 191 fatty1gaf 199  
|:|:|:|  
QY 1 FPFKYAAR 9

RESULT 3  
ID W24686 standard; Protein: 326 AA.  
AC W24686;  
DT 12-AUG-1997 (first entry)  
DE H. pylori inner membrane protein 6093906.aa.  
KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;  
KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;  
KW activator; inhibitor; bacterial life cycle; vaccine; immunise;  
KW detection; antisense; inhibition.  
OS Helicobacter pylori.  
FH Key Location/Qualifiers  
FT misc\_difference 6 /note= "encoded by ARC"  
FT /note= "encoded by ARC"  
PN W09719098-A1.  
PD 29-MAY-1997.  
PE 15-NOV-1996; U18542.  
PR 17-NOV-1995; US-561469.  
PA (ASTR ) ASTRA AB.  
PI Smith DH;  
DR WPI: 97-298052/27.

DR N-PSDB: T77504.  
PT Helicobacter pylori nucleic acid sequences and related proteins -  
PT used for diagnostics and therapeutics  
PS Claim 18; Page 193; 235pp; English.  
CC This sequence represents an H. pylori inner membrane protein.  
CC Helicobacter pylori has been strongly linked to chronic gastritis and  
CC duodenal ulcer disease. The nucleic acid sequences of the invention  
CC are used to evaluate compounds, especially activators or inhibitors of  
CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid  
CC sequence. The nucleic acid sequences, and corresponding proteins, are  
CC also useful for generating vaccines for immunising subjects against H.  
CC pylori or for use in detecting the presence of Helicobacter species in  
CC a sample. Antisense nucleic acid sequences of these sequences are  
CC used to inhibit expression of a gene from Helicobacter species. H.  
CC pylori whole genomic DNA was isolated and nebulised to a median size of  
CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique  
CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are  
CC complementary to the BstXI-cut PMPX vectors, while the overhang is not  
CC self-complementary. Therefore the linkers will not concatamerise nor  
CC will the cut vector re-ligate itself easily. The linker-adaptor inserts  
CC are ligated to each of the 20 PMPX vectors to construct a series of  
CC shotgun subclone libraries. The purified DNA samples were then  
CC sequenced.  
CC Note: The ORF/protein reference number for this sequence was obtained  
CC from the related specification, W09640893.  
SQ Sequence 326 AA;

Query Match 69.4%; Score 50; DB 23; Length 326;  
Best Local Similarity 55.6%; Pred. No. 1.77e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 191 fatty1gaf 199  
|:|:|:|  
QY 1 FPFKYAAR 9

RESULT 4  
ID W20801 standard; protein: 496 AA.  
AC W20801;  
DT 16-JUL-1997 (first entry)  
DE H. pylori inner membrane protein, 09ap11406orf15.  
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
KW binding compound; bacterium; life cycle; activator; bactericidal; inhibitor;  
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.  
OS Helicobacter pylori.  
FH Key Location/Qualifiers  
FT misc\_difference 6 /note= "encoded by ARC"  
FT /note= "encoded by ARC"  
PN W09640893-A1.  
PD 19-DEC-1996.  
PE 06-JUN-1996; U09122.  
PR 07-JUN-1995; US-487032.  
PR 01-APR-1996; US-630405.  
PA (ASTR ) ASTRA AB.  
PI Berglindh OT, Smith D, Mellgaerd BL;  
DR WPI: 97-052306/05.  
DR N-PSDB: T68054.  
PT Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
PS Claim 56; Page 1207-1208; 1481pp; English.  
CC The present sequence is a H. pylori inner membrane protein.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
CC overlapping contigs generated by mechanically shearing the bacterial  
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
SQ Sequence 496 AA;



[illegible]

DR WPI: 94-167383/20.  
 PT New peptide(s) derived from tree pollen allergens - able to induce T cell tolerance, useful in diagnoses and therapy of allergies  
 PS Claim 1; Page 10; 12pp: German.  
 CC The 17kd major allergens from trees of the Order Fagales (esp. birches, hazels and alders) are highly homologous. Peptides derived from the T-cell epitopes of these allergens, partic. from the birch Bet v I allergen are useful for diagnosing tree pollen allergy and for stimulating or blocking T-cells of allergic patients in an allergen-specific manner. The peptides can also be used to provoke tolerance to the allergen-specific T-cells. The peptides have one of the sequences R53560-R53569.  
 CC Sequence 18 AA;

Query Match 68.1%; Score 49; DB 10; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 2.17e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 fpeky 6  
 |||||  
 OY 1 fpeky 5

RESULT 9  
 ID R04605 standard; protein: 160 AA.  
 AC R04605;  
 DT 20-SEP-1990 (first entry)  
 DE Major Birch allergen Bet v I  
 KW major birch allergen; cDNA clone bank; IGF; Betula verrucosa.  
 OS synthetic.  
 FH Key Location/Qualifiers  
 FT modified\_site 83..85  
 FT /label-putative glycosylation site  
 PN WO9004025-A.  
 PD 19-APR-1990.  
 PF 13-JUN-1989; 000058.  
 PR 14-OCT-1988; AF-002554.  
 PA (BIOV-) Bioway Biotechnik P.  
 PI Breitenbach M, Kraft D, Rumpold H, Scheiner O, Breiteneder H, Pellenburger K, Valenta R;  
 DR WPI: 90-147842/19.  
 DR N-PSDB: Q04346.  
 PT Identifying allergen expressing nucleotide sequences in clone bank - by reacting with IGE from allergy patients, and new sequences  
 PT encoding major birch allergen  
 PS Claim 8; Fig 10; 59pp: German.  
 CC The gene encoding this protein is highly homologous with the disease-resistance gene of the pea, which is only expressed in plants with plant pathogens. It is thus expected that upon insertion into plants the allergen gene will cause expression of a resistance protein under conditions of stress or in contact with a pathogen.  
 CC Sequence 160 AA;

Query Match 68.1%; Score 49; DB 1; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 2.17e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 63 fpeky 67  
 |||||  
 OY 1 fpeky 5

RESULT 10  
 ID R21796 standard; protein: 160 AA.  
 AC R21796;  
 DT 08-JUN-1992 (first entry)  
 DE Bet v I allergen of birch.  
 KW Allergic diseases; IGF; tree pollen; Fagales; antipodles; T-cell; response.  
 OS Fagales birch.  
 PN WO9202621-A.  
 PD 20-FEB-1992.

PF 06-MAY-1991; E01479.  
 PR 08-AUG-1990; AF-001688.  
 PR 11-APR-1991; US-683831.  
 PA (BIOV-) BIOWAY BIOTECHN PRO.  
 PI Breiteneder H, Reikertstorfer A, Valenta R, Hoffmann-Sommergruber K;  
 DR WPI: 92-080075/10.  
 DR N-PSDB: Q22008.  
 PT AIn g I, Cor a I and Bet v I allergens - and DNA from alder, hazel and birch, useful in diagnosis or therapy of allergic diseases  
 PS Claim 24; Page 42; 54pp: English.  
 CC The sequence was deduced from the recombinant DNA sequence obtl.  
 CC from ripe pollen birch mRNA. The polypeptide is capable of modifying, in sensitive individuals, an allergic response to pollen of a tree of the order Fagales, by altering the individuals T-cell response. The sequence allows mammals to be tested for allergic reactions to specific tree allergens. The polypeptide CC may be used to challenge the mammal to elicit bronchial, CC conjunctival, dermal, nasal or oral provocation. The polypeptide CC may be used to treat a mammal afflicted with a pollen allergy. CC It is administered in an amount sufficient to hypersensitise the CC mammal to Bet v I.  
 CC See also R21792-802.  
 CC Sequence 160 AA;

Query Match 68.1%; Score 49; DB 4; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 2.17e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 63 fpeky 67  
 |||||  
 OY 1 fpeky 5

RESULT 11  
 ID R53962 standard; protein: 655 AA.  
 AC R53962;  
 DT 06-JAN-1995 (first entry)  
 DE Hepatocyte growth factor converting protease.  
 KW Hepatocyte growth factor; protease; cleavage; active; inactive;  
 KW precursor.  
 OS Homo sapiens.  
 PN EP-596524-A.  
 PD 11-MAY-1994.  
 PF 05-NOV-1993; 117988.  
 PR 05-NOV-1992; JP-296133.  
 PR 20-NOV-1992; JP-312234.  
 PR 20-NOV-1992; JP-312242.  
 PA (SHIM/) SHIMOMURA T.  
 PI (MITU) MITSUBISHI KASEI CORP.  
 PI Kitamura N, Miyazawa K, Morimoto Y, Shimomura T;  
 PI Yamada K;  
 DR WPI: 94-152921/19.  
 DR N-PSDB: Q63951.  
 PT Hepatocyte growth factor converting protease and precursor and PT gene encoding them - for producing active two chain HGF from PT inactive single chain HGF  
 PS Claim 12; Page 21-24; 30pp: English.  
 CC Hepatocyte growth factor converting protease is capable of converting CC inactive single chain hepatocyte growth factor (HGF) into active two CC chain HGF by cleavage at a specific site.  
 CC Sequence 655 AA;

Query Match 66.7%; Score 48; DB 10; Length 655;  
 Best Local Similarity 57.1%; Pred. No. 2.67e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 110 fpeky 99 116  
 |||||  
 OY 1 fpeky 7

|                       |  |                                     |
|-----------------------|--|-------------------------------------|
| RESULT                | 12   |                                     |
| ID                    | R89197   | standard; protein; 655 AA.          |
| AC                    | R89197;  |                                     |
| DT                    | 08-AUG-1996  | (first entry)                       |
| DE                    | Human hepatocellular growth factor single chain precursor protein.       |                                     |
| KW                    | Mature protein; inactive; single chain; hepatocellular growth factor;    |                                     |
| KW                    | liver; human; precursor; dimerisation; double chain; serine protease;    |                                     |
| KW                    | hepatitis.   |                                     |
| OS                    | Homo sapiens.  |                                     |
| FH                    | Key  | Location/Qualifiers                 |
| FT                    | peptide  | 356..655                            |
| FT                    | /note= "mature peptide"  |                                     |
| PN                    | J08027026-A.   |                                     |
| PD                    | 30-JAN-1996  |                                     |
| PR                    | 22-JUL-1994;   | 171207,                             |
| PR                    | 22-JUL-1994;   | JP-171207.                          |
| PA                    | (MITU ) MITSUBISHI CHEM CORP.  |                                     |
| DR                    | UPI; 96-136206/14.   |                                     |
| PT                    | New preventative and therapeutic cpds contg. a 34 kd serine protease     |                                     |
| PT                    | - useful for treatment of liver diseases e.g. hepatitis.                 |                                     |
| PS                    | Claim 4.; Page 6-8; 8pp. Japanese.                                       |                                     |
| CC                    | This is the amino acid sequence of the precursor protein of the inactive |                                     |
| CC                    | single chain form of a hepatocellular growth factor. The mature protein  |                                     |
| CC                    | (R89196) has a mol. wt. of around 34 kd and is derived from the 96 kd    |                                     |
| CC                    | precursor protein. The mature protein corresp. to residues 356-655 of    |                                     |
| CC                    | the precursor protein. The inactive single chain form of the growth      |                                     |
| CC                    | factor is activated by dimerisation of the mature protein. The active    |                                     |
| CC                    | protein is a serine protease which can be used for the treatment of      |                                     |
| SQ                    | Sequence   | 655 AA;                             |
| Query Match           |  | 66.7%; Score 48; DB 17; Length 655; |
| Best Local Similarity |  | 57.1%; Pred. No. 2.67e+02;          |
| Matches               | 4; Conservative  | 3; Mismatches 0; Indels 0; Gaps 0;  |
| Db                    | 110 pfcrfg 116   |                                     |
| Oy                    | 1 PFCKAA 7   |                                     |
| RESULT                | 13   |                                     |
| ID                    | W23830   | standard; Protein; 4302 AA.         |
| AC                    | W23830;  |                                     |
| DT                    | 08-MAY-1998  | (first entry)                       |
| DE                    | Human PKD1 protein.  |                                     |
| KW                    | PKD1; polycystin; polycystic renal degeneration; diagnosis; epitope;     |                                     |
| KW                    | therapy.   |                                     |
| OS                    | Homo sapiens.  |                                     |
| FH                    | Key  | Location/Qualifiers                 |
| FT                    | Region   | 26..270                             |
| FT                    | /note= "contains epitope recognised by a                                 |                                     |
| FT                    | PKD1-specific antibody"  |                                     |
| FT                    | Region   | 26..480                             |
| FT                    | /note= "contains epitope recognised by a                                 |                                     |
| FT                    | PKD1-specific antibody"  |                                     |
| FT                    | Region   | 361..540                            |
| FT                    | /note= "contains epitope recognised by a                                 |                                     |
| FT                    | PKD1-specific antibody"  |                                     |
| FT                    | Region   | 480..700                            |
| FT                    | /note= "contains epitope recognised by a                                 |                                     |
| FT                    | PKD1-specific antibody"  |                                     |
| FT                    | Region   | 541..840                            |
| FT                    | /note= "contains epitope recognised by a                                 |                                     |
| FT                    | PKD1-specific antibody"  |                                     |
| FT                    | Region   | 700..1100                           |
| FT                    | /note= "contains epitope recognised by a                                 |                                     |
| FT                    | PKD1-specific antibody"  |                                     |
| FT                    | Region   | 1011..1220                          |
| FT                    | /note= "contains epitope recognised by a                                 |                                     |
| FT                    | PKD1-specific antibody"  |                                     |
| FT                    | Region   | 2161..2370                          |
| FT                    | /note= "contains epitope recognised by a                                 |                                     |
| FT                    | PKD1-specific antibody"  |                                     |

```

FT FT Region 2723..2931 /note= "contains epitope recognised by a
FT FT /note- PKD1-specific antibody"
FT FT Region 2850..3000 /note= "contains epitope recognised by a
FT FT /note- PKD1-specific antibody"
FT FT Region 2932..3067 /note= "contains epitope recognised by a
FT FT /note- PKD1-specific antibody"
FT FT Region 3100..3280 /note= "contains epitope recognised by a
FT FT /note- PKD1-specific antibody"
FT FT Region 3200..3400 /note= "contains epitope recognised by a
FT FT /note- PKD1-specific antibody"
FT FT Region 3311..3603 /note= "contains epitope recognised by a
FT FT /note- PKD1-specific antibody"
FT FT Region 4090..4302 /note= "contains epitope recognised by a
FT FT /note- PKD1-specific antibody"
FT FT DE19650758-C1.
PN PN 02-JAN-1998.
PD PD 06-DEC-1996. 050758.
PE PE 06-DEC-1996; DE-050758.
PR PR (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA PA Martens R, Veilhagen I, Zentgraf H;
PI WPI: 98-034057/04.
PT PT PKD1 protein fragments, DNA and antibodies - useful for diagnosis
PS PS and therapy of polycystic renal degeneration
PP Disclosure; Page -: 5pp; German.
CC CC This sequence represents a human polycystin, PKD1. This protein is used
CC to generate fragments (see W39170-W39184) which contain epitopes
CC recognised by PKD1-specific antibodies and can be used in the detection
CC and diagnosis of the autosomal dominant condition, polycystic renal
CC degeneration.
SQ Sequence 4302 AA:

Query Match 65.3%; Score 47; DB 28; Length 4302;
Best Local Similarity 44.4%; Pred. No. 3.27e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0.

Db 1408 fpyrytwdf 1416
QY 1 PPFKRYAAAF 9

RESULT 14
ID W33396 standard; Protein: 4302 AA.
AC W33396;
DT 01-JUN-1998 (first entry)
DE Human PKD1 polypeptide.
KW Human; polycystic kidney disease 1; PKD1; treatment;
KW autosomal dominant polycystic kidney disease; APKD.
OS Homo sapiens.
PN WO9744457-A1.
PD 27-NOV-1997.
PE 22-MAY-1997; U08799.
PR 03-JUN-1996; US-658136.
RA 24-MAY-1996; US-655360.
PA (GENZ ) GENZYME CORP.
PI Bun T, Connors T, Dackowski W, Germino G, Kilinger K,
PI Qian F;
PI WPI: 98-018511/02.
DR N-PSDB: T94012.
PT Human polycystic kidney disease gene, PKD1 - useful to treat and
PT diagnose human autosomal or adult onset polycystic kidney disease
PS Claim 8; Pages 119-138; 257pp; English.
CC The present sequence is the human polycystic kidney disease 1
CC (PKD1) polypeptide. The PKD1 cDNA or polypeptide may be used to
CC treat autosomal dominant polycystic kidney disease (APKD), and
CC identify carriers of mutant PKD1 genes, i.e. subjects susceptible
CC to APKD. Antibodies (Ab) that distinguish between normal and mutant

```

CC PKD1 sequences can also be used in diagnostic tests. Anti-PKD1 Ab  
CC may also be used to perform subcellular and histochemical  
CC localisation studies, and to block the function of PKD1. Ab are  
CC also useful in rational drug design studies to identify and test  
CC inhibitors of PKD1. Sense and antisense sequences derived from the  
CC PKD1 gene may used for detection and therapy.  
SQ Sequence 4302 AA;

Query Match 65.38; Score 47; DB 29; Length 4302;  
Best Local Similarity 44.48; Pred. No. 3.27e+02;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1408 fpyrywdf 1416  
11:1:1:1  
Qy 1 fpekyvAAF 9

RESULT 15  
ID W00870 standard; Protein; 4302 AA.  
AC W00870;  
DT 02-FEB-1997 (first entry)  
DE Polycystic kidney disease 1 (PKD1) polypeptide.  
KW Adult polycystic kidney disease; APKD; PKD1 gene; diagnosis;  
KW therapy; polycystin.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT peptide 1..23  
FT /label= Sig\_peptide  
FT 24..4303  
FT /label= Mat\_protein  
FT region 72..125  
FT /label= LRR  
FT /note= "leucine-rich repeat region"  
FT domain 2580..2600  
FT /label= TM1  
FT /note= "transmembrane domain 1"  
FT 2693..2713  
FT /label= TM2  
FT /note= "transmembrane domain 2"  
FT 3075..3095  
FT /label= TM3  
FT /note= "transmembrane domain 3"  
FT 3281..3301  
FT /label= TM4  
FT /note= "transmembrane domain 4"  
FT 3323..3343  
FT /label= TM5  
FT /note= "transmembrane domain 5"  
FT 3559..3579  
FT /label= TM6  
FT /note= "transmembrane domain 6"  
FT 3582..3612  
FT /label= TM7  
FT /note= "transmembrane domain 7"  
FT 3669..3689  
FT /label= TM8  
FT /note= "transmembrane domain 8"  
FT misc\_difference 50  
FT /label= N-glycosylation\_site  
FT 89  
FT /label= N-glycosylation\_site  
FT 107  
FT /label= N-glycosylation\_site  
FT 112  
FT /label= N-glycosylation\_site  
FT 187  
FT /label= N-glycosylation\_site  
FT misc\_difference 621  
FT /label= N-glycosylation\_site  
FT 632  
FT /label= N-glycosylation\_site  
FT 746  
FT /label= N-glycosylation\_site

FT misc\_difference 810  
FT /label= N-glycosylation\_site  
FT 841  
FT /label= N-glycosylation\_site  
FT 854  
FT /label= N-glycosylation\_site  
FT 890  
FT /label= N-glycosylation\_site  
FT 921  
FT /label= N-glycosylation\_site  
FT 1004  
FT /label= N-glycosylation\_site  
FT 1034  
FT /label= N-glycosylation\_site  
FT 1072  
FT /label= N-glycosylation\_site  
FT 1113  
FT /label= N-glycosylation\_site  
FT 1178  
FT /label= N-glycosylation\_site  
FT 1194  
FT /label= N-glycosylation\_site  
FT 1240  
FT /label= N-glycosylation\_site  
FT 1269  
FT /label= N-glycosylation\_site  
FT 1336  
FT /label= N-glycosylation\_site  
FT 1348  
FT /label= N-glycosylation\_site  
FT 1382  
FT /label= N-glycosylation\_site  
FT 1450  
FT /label= N-glycosylation\_site  
FT 1455  
FT /label= N-glycosylation\_site  
FT 1474  
FT /label= N-glycosylation\_site  
FT 1518  
FT /label= N-glycosylation\_site  
FT 1541  
FT /label= N-glycosylation\_site  
FT 1554  
FT /label= N-glycosylation\_site  
FT 1563  
FT /label= N-glycosylation\_site  
FT 1647  
FT /label= N-glycosylation\_site  
FT 1661  
FT /label= N-glycosylation\_site  
FT 1733  
FT /label= N-glycosylation\_site  
FT 1791  
FT /label= N-glycosylation\_site  
FT 1834  
FT /label= N-glycosylation\_site  
FT 1867  
FT /label= N-glycosylation\_site  
FT 1880  
FT /label= N-glycosylation\_site  
FT 1991  
FT /label= N-glycosylation\_site  
FT 2050  
FT /label= N-glycosylation\_site  
FT 2074  
FT /label= N-glycosylation\_site  
FT 2125  
FT /label= N-glycosylation\_site  
FT 2248  
FT /label= N-glycosylation\_site  
FT 2353  
FT /label= N-glycosylation\_site  
FT 2395



**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
 WISE (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Msrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Sep 11 12:37:54 1998; Maspar time 3.45 Seconds  
 Tabular output not generated. 95.378 Million cell updates/sec

Title: >US-08-452-843-1  
 Description: (1-9) from US08452843.pep  
 Perfect Score: 72  
 Sequence: 1 FPFKYAAF 9

Scoring table: PAM 150  
 Gap 15

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: p156  
 1:p1r1 2:p1r2 3:p1r3 4:p1r4 5:nr13d

Statistics: Mean 23.322; Variance 37.313; scale 0.625

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description | Pred. No.             |
|------------|-------|-------------|-----------|----|-------------|-----------------------|
| 1          | 55    | 76.4        | 173       | 2  | B69143      | hypothetical protein  |
| 2          | 55    | 76.4        | 469       | 2  | S74825      | hypothetical protein  |
| 3          | 54    | 75.0        | 2178      | 2  | S55805      | alpha-toxin - Clostri |
| 4          | 52    | 72.2        | 503       | 2  | S73843      | general amino acid pe |
| 5          | 52    | 72.2        | 529       | 2  | B49993      | glycylpeptide N-tetra |
| 6          | 52    | 72.2        | 702       | 2  | A34434      | arylphorin alpha ch   |
| 7          | 51    | 70.8        | 211       | 2  | E64710      | hypothetical protein  |
| 8          | 51    | 70.8        | 404       | 2  | S45923      | probable phosphopane  |
| 9          | 51    | 70.8        | 593       | 2  | S16375      | surface-layer glycopr |
| 10         | 51    | 70.8        | 593       | 2  | S16375      | surface-layer glycopr |
| 11         | 50    | 69.4        | 88        | 2  | D70010      | hypothetical protein  |
| 12         | 50    | 69.4        | 132       | 2  | PC2131      | hypothetical protein  |
| 13         | 50    | 69.4        | 451       | 2  | A38099      | hepatocyte growth fac |
| 14         | 50    | 69.4        | 490       | 2  | A64679      | glycylpeptide N-tetra |
| 15         | 50    | 69.4        | 549       | 2  | B65215      | hypothetical 59.2 kD  |
| 16         | 50    | 69.4        | 589       | 2  | A34341      | MADH-ubiquitinone ox  |
| 17         | 50    | 69.4        | 664       | 2  | S73624      | poly(3-hydroxybutyrat |
| 18         | 50    | 69.4        | 840       | 2  | A42970      | MG366 homolog G12-ori |
| 19         | 50    | 69.4        | 890       | 2  | S54554      | H+-transporting ATPas |
| 20         | 50    | 69.4        | 1004      | 2  | A48821      | H+-transporting ATPas |
| 21         | 50    | 69.4        | 1379      | 2  | S01234      | mnt-5 protein - fruit |
| 22         | 50    | 69.4        | 1390      | 1  | TVHUME      | hepatocyte growth fac |
| 23         | 49    | 68.1        | 160       | 2  | E55699      | hepatocyte growth fac |
|            |       |             |           |    |             | major pollen allergen |

| 24 | 49 | 68.1 | 160  | 2 | S41901 | gene Betv1n protein - |
|----|----|------|------|---|--------|-----------------------|
| 25 | 49 | 68.1 | 160  | 2 | S05376 | major pollen allergen |
| 26 | 49 | 68.1 | 160  | 2 | G55699 | major pollen allergen |
| 27 | 49 | 68.1 | 160  | 2 | C55699 | major pollen allergen |
| 28 | 49 | 68.1 | 160  | 2 | S41905 | gene Betv1n protein - |
| 29 | 49 | 68.1 | 160  | 2 | F55699 | major pollen allergen |
| 30 | 49 | 68.1 | 160  | 2 | I55699 | major pollen allergen |
| 31 | 49 | 68.1 | 291  | 2 | H69521 | 4-Hydroxybenzoate oct |
| 32 | 49 | 68.1 | 355  | 2 | A32115 | staro91-COA desatur   |
| 33 | 49 | 68.1 | 391  | 2 | S72717 | Leb1170_F3_112 prote  |
| 34 | 49 | 68.1 | 427  | 2 | S73659 | MG288 homolog P02_ori |
| 35 | 49 | 68.1 | 545  | 2 | B44054 | oriz protein - Junoni |
| 36 | 49 | 68.1 | 607  | 2 | C40361 | vicr-region hypothe   |
| 37 | 49 | 68.1 | 748  | 2 | A45243 | envelope protein Hrp  |
| 38 | 49 | 68.1 | 822  | 2 | H69547 | molybdopterin oxidore |
| 39 | 49 | 68.1 | 1342 | 2 | A31946 | xanthine dehydrogenas |
| 40 | 49 | 68.1 | 2470 | 2 | I50726 | cation-independent ma |
| 41 | 48 | 66.7 | 585  | 2 | S77114 | hypothetical protein  |
| 42 | 48 | 66.7 | 775  | 2 | E69425 | hepatocyte growth fac |
| 43 | 48 | 66.7 | 956  | 2 | A65072 | hypothetical protein  |
| 44 | 48 | 66.7 | 1353 | 2 | J00407 | hypothetical protein  |
| 45 | 48 | 66.7 | 1353 | 2 | J00407 | xanthine dehydrogenas |

## ALIGNMENTS

| RESULT ENTRY          | 1   | ALIGNMENTS     |
|-----------------------|---|----------------|
| TITLE                 | B69143  | #type complete |
| ORGANISM              | hypothetical protein MTH336 - Methanobacterium              |                |
| DATE                  | thermoautotrophicum (strain Delta H)                        |                |
| ACCESSIONS            | #formal_name Methanobacterium thermoautotrophicum           |                |
| REFERENCE             | 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change     |                |
| #authors              | B69143  |                |
| #journal              | Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; |                |
| #title                | Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;     |                |
| #cross-references     | Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; |                |
| #accession            | Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vitale, R.;  |                |
| #status               | Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso,  |                |
| #molecule_type        | A.; Bush, D.; Sifer, H.; Patwell, D.; Prabhakar, S.;        |                |
| #residues             | McDougal, S.; Shiner, G.; Goyal, A.; Pietrovski, S.;        |                |
| GENETICS              | Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,   |                |
| ##experimental_source | J.; Reeve, J.N.   |                |
| ##strain              | J. Bacteriol. (1997) 179:7135-7155                          |                |
| ##gene                | Complete genome sequence of Methanobacterium                |                |
| ##codon               | thermoautotrophicum Delta H: functional analysis and        |                |
| ##start_codon         | comparative genomics.                                       |                |
| ##length              | #accession B69143   |                |
| ##length              | preliminary: nucleic acid sequence not shown;               |                |
| ##length              | translation not shown                                       |                |
| ##length              | ##molecule_type DNA   |                |
| ##length              | 1-173 #label MTH  |                |
| ##length              | ##cross-references GB:AE000666                              |                |
| ##length              | ##experimental_source strain Delta H                        |                |
| ##length              | ##gene MTH336   |                |
| ##length              | ##codon TTG   |                |
| ##length              | ##length 173 #molecular_weight 20426 #checksum 9042         |                |
| ##length              | Query Match   |                |
| ##length              | Best Local Similarity 55.6%; Pred. No. 5.80e+00;            |                |
| ##length              | Matches 5; Conservative 2; Mismatches 2; Gaps 0;            |                |
| DB                    | 32 FPFKYAAF 40  |                |
| Qy                    | 1 FPFKYAAF 9  |                |
| RESULT ENTRY          | 2   |                |
| TITLE                 | S74825  | #type complete |
|                       | hypothetical protein slr1747 - Synechocystis sp. (PCC 6803) |                |

```

ORGANISM      #formal_name Synechocystis sp.
DATE           #variety PCC 6803
               #sequence_revision 25-Apr-1997 #text_change
               30-Jan-1998
ACCESSIONS     S74825
REFERENCE      #authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
               Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugitara, M.;
               Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
               Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
               S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
               Yasuda, M.; Tabata, S.
               #journal DNA Res. (1996) 3:109-116
               #title Sequence analysis of the genome of the unicellular
               cyanobacterium Synechocystis sp. PCC6803. II. Sequence
               determination of the entire genome and assignment of
               potential protein-coding regions.
               #cross-references EMBL:Z48636; NID:g728537; PID:g755724
               #accession S74825
               #status nucleic acid sequence not shown; translation not shown
               #molecule_type DNA
               #residues 1-469 #label KAN
               #cross-references EMBL:D90909; NID:g1652844; PID:d1018519; PID:g1652868
               #note the nucleotide sequence was submitted to the EMBL Data
               Library, June 1996
SUMMARY        #length 469 #molecular_weight 52543 #checksum 1395
Query Match    76.4%; Score 55; DB 2; Length 469;
Best Local Similarity 55.6%; Pred. No. 5.80e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db            282 FPFKXAAAF 9
QY            1 FPFKXAAAF 9

RESULT 3
ENTRY      S55805 #type complete
TITLE      alpha-toxin - Clostridium novyi (ATCC 19402)
ORGANISM   #formal_name Clostridium novyi
            #variety ATCC 19402
DATE       28-Oct-1996 #sequence_revision 08-Nov-1996 #text_change
            10-Sep-1997
ACCESSIONS S55805; S71294; S71158; S44273; I40834; S44272
REFERENCE   #authors Hofmann, F.; Hermann, A.; Habermann, E.; von
            Eichel-Streiber, C.
            #journal Mol. Gen. Genet. (1995) 247:670-679
            #title Sequencing and analysis of the gene encoding the alpha-toxin
            of Clostridium novyi proves its homology to toxins A and B
            of Clostridium difficile.
            #cross-references EMBL:Z5342160
            #accession S55805
            #status nucleic acid sequence not shown
            #molecule_type DNA
            #residues 1-2178 #label HOF
            #cross-references EMBL:Z48636; NID:g728537; PID:g755724
            #accession S71294
            #molecule_type protein
            #residues 1-15 #label HOW
REFERENCE     S71158
#authors      Hofmann, F.
#submission   submitted to the EMBL Data Library, March 1995
#accession    S71158
#molecule_type DNA
#residues     1-1179, 'LKV', 1183, 'LYTHICE', 1191-2178 #label HOS
#cross-references EMBL:Z48636; NID:g728537; PID:g755724
#accession    S44272
#authors      Hofmann, F.; Habermann, E.; von Eichel-Streiber, C.
#submission   submitted to the EMBL Data Library, July 1993
#description   Sequence analysis of Clostridium novyi alpha-toxin: a member
               of the family of large clostridial cytotoxins.
#accession    S44273

```

```

#molecule_type DNA
#residues 1-243;1204-2178 #label HOA
#cross-references EMBL:Z23281
GENETICS
#gene       tcn-alpha
CLASSIFICATION #superfamily cpl repeat homology
KEYWORDS      virulence factor
FEATURE       1880-1899
SUMMARY       #domain cpl repeat homology #label cpl2
               #length 2178 #molecular_weight 250166 #checksum 5975
Query Match    75.0%; Score 54; DB 2; Length 2178;
Best Local Similarity 66.7%; Pred. No. 8.54e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db            1273 FPFKXAAAF 1281
QY            1 FPFKXAAAF 9

RESULT 4
ENTRY      S73843 #type complete
TITLE      general amino acid permease Gap1 homolog F10_orf503 -
            Mycoplasma pneumoniae (ATCC 29342) (SGC3)
ALTERNATE_NAMES hypothetical protein F10_orf503
ORGANISM     #formal_name Mycoplasma pneumoniae
            #variety ATCC 29342
DATE       27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change
            09-Sep-1997
ACCESSIONS S73843
REFERENCE   S73327
#authors    Himmelfreuch, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li,
            B.C.; Hermann, R.
            #journal Nucleic Acids Res. (1996) 24:4420-4449
            #title Complete sequence analysis of the genome of the bacterium
            Mycoplasma pneumoniae.
            #accession S73843
            #status preliminary; nucleic acid sequence not shown;
            translation not shown
            #molecule_type DNA
            #residues 1-503 #label HIM
            #cross-references EMBL:AE000051; NID:g1674211; PID:g1674212
            #note the nucleotide sequence was submitted to the EMBL Data
            Library, November 1996
GENETICS
#gene       gap1
#genetic_code SGC3
SUMMARY     #length 503 #molecular_weight 54960 #checksum 5162
Query Match    72.2%; Score 52; DB 2; Length 503;
Best Local Similarity 55.6%; Pred. No. 1.83e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db            23 FAFNYVAGF 31
QY            1 FPFKXAAAF 9

RESULT 5
ENTRY      B49993 #type complete
TITLE      glycy1peptide N-tetradecanoyltransferase (EC 2.3.1.97) -
            Ajellomyces capsulata
ORGANISM     #formal_name Ajellomyces capsulata, Histoplasma capsulatum
            10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change
            09-Sep-1997
ACCESSIONS B49993
REFERENCE   B49993
#authors     Lodge, J.K.; Johnson, R.L.; Weinberg, R.A.; Gordon, J.I.
            #journal J. Biol. Chem. (1994) 269:2996-3009
            #title Comparison of myristoyl-CoA:protein N-myristoyltransferases
            from three pathogenic fungi: Cryptococcus neoformans,
            Histoplasma capsulatum, and Candida albicans.
#accession    B49993

```



```
##status Preliminary
##molecule_type DNA
##residues 1-529 ##label LOD
##cross-references GB:L25118; NID:g407694; PID:g407695

GENETICS
#gene Nmt
#introns 203/2; 464/3
KEYWORDS acyltransferase
SUMMARY #length 529 #molecular-weight 59363 #checksum 3672

Query Match 72.2%; Score 52; DB 2; Length 529;
Best Local Similarity 55.6%; Pred. No. 1.83e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 192 FRFVSPAF 200
|:|:|:|
QY 1 FPFKYAAAF 9

RESULT 6
ENTRY A34434 #type complete
TITLE arylphorin alpha chain precursor - tobacco hornworm
ORGANISM #formal_name Manduca sexta #common_name tobacco hornworm
DATE 15-Jun-1990 #sequence_revision 15-Jun-1990 #text_change 31-Oct-1997

ACCESSIONS A34434
REFERENCE A34434
#authors Willott, E.; Wang, X.Y.; Wells, M.A.
#journal J. Biol. Chem. (1989) 264:19052-19059
#title cDNA and gene sequence of Manduca sexta arylphorin, an aromatic amino acid-rich larval serum protein. Homology to arthropod hemocyanins.
#cross-references M01D:90037032
#accession A34434
#status Preliminary
#molecule_type DNA
#residues 1-702 ##label WIL
##cross-references GB:M28396; EMBL:J05092; NID:g159486; PID:g159487; EMBL:J05093

SUMMARY #length 702 #molecular-weight 83866 #checksum 4883

Query Match 72.2%; Score 52; DB 2; Length 702;
Best Local Similarity 44.4%; Pred. No. 1.83e+01;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 686 FPKYKVPF 694
|:|:|:|
QY 1 FPFKYAAAF 9

RESULT 7
ENTRY E64710 #type complete
TITLE hypothetical protein HP1525 - Helicobacter pylori (strain 26695)
ORGANISM #formal_name Helicobacter pylori
DATE 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 10-Oct-1997
E64710
A64520
REFERENCE A64520
#authors Tomb, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khakhria, H.G.; Glodek, A.; McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.R.; Berg, D.E.; Gocayne, J.D.; Uitterlinden, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Matthey, L.; Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C.M.; Venter, J.C.
#journal Nature (1997) 388:539-547
#title The complete genome sequence of the gastric pathogen Helicobacter pylori.

##cross-references M01D:97394467
#accession E64710
#status Preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-211 ##label TOM
##cross-references GB:AE000650; GB:AE000511; NID:g2314700; PID:g2314707; TIGR:HP1525

GENETICS
#start_codon GTG
SUMMARY #length 211 #molecular-weight 24866 #checksum 4606

Query Match 70.8%; Score 51; DB 2; Length 211;
Best Local Similarity 55.6%; Pred. No. 2.65e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 189 FAFYDSAF 197
|:|:|:|
QY 1 FPFKYAAAF 9

RESULT 8
ENTRY S45923 #type complete
TITLE probable phosphopantetheine-binding protein - yeast (Saccharomyces cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 14-Nov-1997

ACCESSIONS S45923
REFERENCE S45816
#authors Dondey, H.; Gassenhuber, H.; Obermaier, B.; Piravandi, E.
#submission submitted to the Protein Sequence Database, August 1994
#accession S45923
#molecule_type DNA
#residues 1-404 ##label DOM
##cross-references EMBL:Z35932; NID:g536306; PID:g536307; MIPS:YBR063C
##experimental_source strain S288C

GENETICS
#map_position 2R
CLASSIFICATION #superfamily acyl carrier protein homology
KEYWORDS phosphopantetheine; transmembrane protein
FEATURE 93-113
#domain transmembrane #status predicted #label TM\
#binding_site phosphopantetheine (Ser) (covalent)
#status predicted

SUMMARY #length 404 #molecular-weight 46444 #checksum 1399

Query Match 70.8%; Score 51; DB 2; Length 404;
Best Local Similarity 44.4%; Pred. No. 2.65e+01;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 10 FPYEGSDF 18
|:|:|:|
QY 1 FPFKYAAAF 9

RESULT 9
ENTRY S16375 #type complete
TITLE surface-layer glycoprotein precursor - Methanothermobacter sociabilis
ORGANISM #formal_name Methanothermobacter sociabilis
DATE 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 09-Sep-1997
S16375; S26144; S26098; S21874
S16225
REFERENCE S16225
#authors Broeckl, G.; Behr, M.; Fabry, S.; Hensel, R.; Kaudewitz, H.; Biedl, E.; Koenig, H.
#journal Eur. J. Biochem. (1991) 199:147-152
#title Analysis and nucleotide sequence of the genes encoding the surface-layer glycoproteins of the hyperthermophilic methanogen Methanothermobacter feridus and Methanothermobacter sociabilis.
```

```

#cross-references MUID:91293115
#accession S16375
#ENTRY
#molecule-type DNA
#residues 1-593 ##label BRO
#cross-references EMBL:X58296
#accession S26144
#molecule-type protein
#residues 23-42 ##label BRO2
REFERENCE S21873
#authors Broeckl, G.
#submission submitted to the EMBL Data Library, March 1991
#accession S26098
#molecule-type DNA
#residues 1-256, 'I', '258', 'V', '260-593 ##label BRO1
#cross-references EMBL:X58296; NID:944546; PID:9809717
GENETICS
#gene sigA
#start-codon GTG
KEYWORDS glycoprotein
FEATURE
1-22
23-593
#domain signal sequence #status experimental #label SIG\
#product surface-layer glycoprotein #status experimental
#label MAT
SUMMARY #length 593 #molecular-weight 65503 #checksum 8058

Query Match 70.8%; Score 51; DB 2; Length 593;
Best Local Similarity 62.5%; Pred. No. 2.65e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 526 YPFKYAVS 533
OY 1 PPFKYAAA 8

RESULT 10
ENTRY S16225 #type complete
TITLE surface-layer glycoprotein precursor - Methanothermus
ORGANISM #formal_name Methanothermus fervidus
DATE 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change
ACCESSIONS S16225; S21873
REFERENCE S16225
#authors Broeckl, G.; Behr, M.; Fabry, S.; Hensel, R.; Kaudewitz, H.;
#journal Eur. J. Biochem. (1991) 199:147-152
#title Analysis and nucleotide sequence of the genes encoding the
#molecule-type DNA surface-layer glycoproteins of the hyperthermophilic
#residues 1-593 ##label BRO methanogens Methanothermus fervidus and Methanothermus
#cross-references MUID:91293115 sociabilis.
#accession S16225
#molecule-type DNA
#residues 1-593 ##label BRO
#cross-references EMBL:X58297; NID:944281; PID:9809714
GENETICS
#gene sigA
#start-codon GTG
KEYWORDS glycoprotein
FEATURE
1-22
23-593
#domain signal sequence #status predicted #label SIG\
#product surface-layer glycoprotein #status predicted
#label MAT
SUMMARY #length 593 #molecular-weight 65481 #checksum 7421

Query Match 70.8%; Score 51; DB 2; Length 593;
Best Local Similarity 62.5%; Pred. No. 2.65e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 526 YPFKYAVS 533
OY 1 PPFKYAAA 8

```

```

RESULT 11
ENTRY D70010 #type complete
TITLE hypothetical protein yugF - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSIONS D70010
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Allion, G.; Azevedo, V.; Bertero, M.G.; Bessières, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.R.; Codani, J.J.; Connerthon, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Dueserhoef, A.;
Ehrlich, S.D.; Emerson, P.T.; Enlitan, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Funa, S.; Gallizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarovic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Noesti, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Prescan, E.; Pujic, P.; Purrelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadate, I.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scorfione, E.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serrier, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.;
Tampstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambolt, R.; Wedler, E.; Wedler, H.; Weltenegeger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
#accession D70010 Bacillus subtilis.
#status preliminary; nucleic acid sequence not shown;
#molecule-type DNA translation not shown
#residues 1-88 ##label KUN
#experimental_source strain 168
GENETICS
#gene yugF
#start-codon ATG
SUMMARY #length 88 #molecular-weight 10065 #checksum 2766

Query Match 69.4%; Score 50; DB 2; Length 88;
Best Local Similarity 62.5%; Pred. No. 3.84e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 21 PFKYGEER 28
OY 2 PFKYAAAF 9

RESULT 12
ENTRY PC2131 #type fragment
TITLE hepatocyte growth factor receptor - rat (fragment)
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 03-May-1994 #sequence_revision 07-Oct-1994 #text_change
ACCESSIONS PC2131

```

```

REFERENCE
#authors PC2131
Tsujii, M.; Kawano, S.; Tsujii, S.; Ito, T.; Hayashi, N.;
Horimoto, M.; Mita, E.; Nagano, K.; Masuda, E.; Hayashi,
N.; Fusamoto, H.; Kamada, T.
#journal Biochem. Biophys. Res. Commun. (1994) 200:536-541
#title Increased expression of c-met messenger RNA following acute
gastric injury in rats.
#accession PC2131
#molecule_type mRNA
#residues 1-132 #label TSU
COMMENT This protein participates in the healing process of gastric mucosa
after injury.
GENETICS
#gene C-met
#superfamily heptocyte growth factor receptor; protein
CLASSIFICATION
#kinase homology
KEYWORDS
SUMMARY
#length 132 #checksum 7223
Query Match 69.4%; Score 50; DB 2; Length 132;
Best Local Similarity 55.6%; Pred. No. 3.84e+01;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 3 EPIKYNDF 11
11:111
1
QY 1 PPFKYAAAF 9

RESULT 13
ENTRY A38099 #type complete
TITLE glycyloleptide N-tetradecanoyltransferase (EC 2.3.1.97) -
#organism Yeast (Candida albicans)
#formal_name Candida albicans
#organism 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
DATE 09-Sep-1997
ACCESSIONS A38099
REFERENCE A38099
#authors Wiegand, R.C.; Carr, C.; Minnerly, J.C.; Pauley, A.M.;
Carroll, C.P.; Langner, C.A.; Duronio, R.J.; Gordon, J.I.
#journal J. Biol. Chem. (1992) 267:8591-8598
#title The Candida albicans myristoyl-CoA:protein
N-myristoyltransferase gene. Isolation and expression in
Saccharomyces cerevisiae and Escherichia coli.
#accession A38099
#status preliminary; not compared with conceptual translation
#molecule_type DNA
#residues 1-451 #label WIE
#cross-references GB:M80544; NID:g170883; PID:g170884
KEYWORDS acyltransferase
SUMMARY #length 451 #molecular-weight 51877 #checksum 1063
Query Match 69.4%; Score 50; DB 2; Length 451;
Best Local Similarity 55.6%; Pred. No. 3.84e+01;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 115 PFKYSHEF 123
1111111
1
QY 1 PPFKYAAAF 9

RESULT 14
ENTRY A64679 #type complete
TITLE NADH-ubiquinone oxidoreductase, NQO14 subunit - Helicobacter
pylori (strain 26695)
#organism #formal_name Helicobacter pylori
DATE 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
10-Oct-1997
ACCESSIONS A64679
REFERENCE A64679
#authors Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klein,
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;

```

```

Richardson, D.; Dodson, R.; Khalak, H.G.; Glodex, A.;
McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Uterback, T.R.;
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Welden, J.M.;
Fuji, C.; Bowman, C.; Matthey, L.; Wallin, E.; Hayes,
W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser,
C.M.; Venter, J.C.
#journal Nature (1997) 388:539-547
#title The complete genome sequence of the gastric pathogen
Helicobacter pylori.
#cross-references GB:AE000631; GB:AE000511; NID:g2314421; PID:g2314438;
#accession A64679
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-490 #label TOM
#cross-references GB:AE000631; GB:AE000511; NID:g2314421; PID:g2314438;
TIGR:HP1273
SUMMARY #length 490 #molecular-weight 55164 #checksum 5890
Query Match 69.4%; Score 50; DB 2; Length 490;
Best Local Similarity 55.6%; Pred. No. 3.84e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 339 PAFYIGAF 347
1111111
1
QY 1 PPFKYAAAF 9

RESULT 15
ENTRY B65215 #type complete
TITLE hypothetical 58.2 kD protein in soxR-acs intergenic region -
#organism Escherichia coli (strain K-12)
#formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
14-Nov-1997
ACCESSIONS B65215
REFERENCE B65215
#authors Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession B65215
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-549 #label BLAT
#cross-references GB:AE000480; GB:U00096; NID:g2367344; PID:g1790503;
UMCP:b4067
SUMMARY #experimental_source strain K-12, substrain MG1655
GENETICS
#gene yjcG
#length 549 #molecular-weight 59197 #checksum 5615
Query Match 69.4%; Score 50; DB 2; Length 549;
Best Local Similarity 55.6%; Pred. No. 3.84e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 494 PFRYPALF 502
1111111
1
QY 1 PPFKYAAAF 9

Search completed: Fri Sep 11 12:38:25 1998
Job time : 31 secs.

```

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
WISCONSIN  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_pp protein - protein database search, using Smith-Waterman algorithm.

Run on: Fri Sep 11 12:38:43 1998; Maspar time 2.48 Seconds  
91.171 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-1  
Description: (1-9) from US08452843.pep  
Perfect Score: 72  
Sequence: 1 FPFKYAAAF 9

Scoring table: PAM 150  
Gap 15

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swiss1

Statistics: Mean 24.345; Variance 32.387; scale 0.752

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result | Query | Match | Length | DB | ID          | Description            | Pred. No. |
|--------|-------|-------|--------|----|-------------|------------------------|-----------|
| No.    | Score |       |        |    |             |                        |           |
| 1      | 52    | 72.2  | 503    | 1  | Y226_MYCPN  | HYPOTHETICAL PROTEIN M | 5.26e+00  |
| 2      | 52    | 72.2  | 529    | 1  | NMT_AJECA   | GLYCILPEPTIDE N-TETRAD | 5.26e+00  |
| 3      | 52    | 72.2  | 702    | 1  | ARYA_YANSE  | ARYLPHORIN ALPHA SUBUN | 5.26e+00  |
| 4      | 51    | 70.8  | 404    | 1  | YBR3_YEAST  | HYPOTHETICAL 46.4 KD P | 8.09e+00  |
| 5      | 51    | 70.8  | 593    | 1  | CSG_METFE   | CELL SURFACE GLYCOPROT | 8.09e+00  |
| 6      | 51    | 70.8  | 593    | 1  | CSG_METSC   | CELL SURFACE GLYCOPROT | 8.09e+00  |
| 7      | 50    | 69.4  | 451    | 1  | NMT_CANAL   | GLYCILPEPTIDE N-TETRAD | 1.24e+01  |
| 8      | 50    | 69.4  | 549    | 1  | YJCG_ECOLI  | HYPOTHETICAL 58.2 KD P | 1.24e+01  |
| 9      | 50    | 69.4  | 589    | 1  | PHBC_ALCEU  | POLY-BETA-HYDROXYBUTYR | 1.24e+01  |
| 10     | 50    | 69.4  | 664    | 1  | Y366_MYCPN  | HYPOTHETICAL PROTEIN M | 1.24e+01  |
| 11     | 50    | 69.4  | 840    | 1  | VPH1_YEAST  | VACUOLAR APP SYNTHASE  | 1.24e+01  |
| 12     | 50    | 69.4  | 890    | 1  | STVL_YEAST  | PROTEIN DMMT-5 PRECURS | 1.24e+01  |
| 13     | 50    | 69.4  | 1010   | 1  | NMT5_DROME  | HEPATOCYTE GROWTH FACT | 1.24e+01  |
| 14     | 50    | 69.4  | 1379   | 1  | MET_MOUSE   | HEPATOCYTE GROWTH FACT | 1.24e+01  |
| 15     | 50    | 69.4  | 1390   | 1  | MET_HUMAN   | HEPATOCYTE GROWTH FACT | 1.24e+01  |
| 16     | 49    | 68.1  | 131    | 1  | MCRD_METFE  | METHYL-COENZYME M REDU | 1.87e+01  |
| 17     | 49    | 68.1  | 159    | 1  | BVIG_BETVE  | MAJOR POLLEN ALLERGEN  | 1.87e+01  |
| 18     | 49    | 68.1  | 159    | 1  | BVIG_BETVE  | MAJOR POLLEN ALLERGEN  | 1.87e+01  |
| 19     | 49    | 68.1  | 159    | 1  | BVIG_BETVE  | MAJOR POLLEN ALLERGEN  | 1.87e+01  |
| 20     | 49    | 68.1  | 159    | 1  | BVIG_BETVE  | MAJOR POLLEN ALLERGEN  | 1.87e+01  |
| 21     | 49    | 68.1  | 159    | 1  | BVIG_BETVE  | MAJOR POLLEN ALLERGEN  | 1.87e+01  |
| 22     | 49    | 68.1  | 159    | 1  | BVIG_BETVE  | MAJOR POLLEN ALLERGEN  | 1.87e+01  |
| 23     | 49    | 68.1  | 308    | 1  | IFRH_SOLUTU | ISOFLAVONE REDUCTASE H | 1.87e+01  |

|    |    |      |      |   |            |                        |          |
|----|----|------|------|---|------------|------------------------|----------|
| 24 | 49 | 68.1 | 355  | 1 | ACOL_MOUSE | ACYL-COA DESATURASE 1  | 1.87e+01 |
| 25 | 49 | 68.1 | 427  | 1 | Y96E_MYCPN | HYPOTHETICAL PROTEIN M | 1.87e+01 |
| 26 | 49 | 68.1 | 450  | 1 | NMT_CABEL  | PROBABLE GLYCILPEPTIDE | 1.87e+01 |
| 27 | 49 | 68.1 | 607  | 1 | YSCC_YEREN | YOP PROTEINUS TRANSLOC | 1.87e+01 |
| 28 | 49 | 68.1 | 735  | 1 | YDDB_SCHPO | HYPOTHETICAL 83.3 KD P | 1.87e+01 |
| 29 | 49 | 68.1 | 748  | 1 | HRPH_PSEST | HYPERSENSITIVITY RESPO | 1.87e+01 |
| 30 | 49 | 68.1 | 1344 | 1 | XDR_DROPS  | XANTHINE DEHYDROGENASE | 1.87e+01 |
| 31 | 49 | 68.1 | 1344 | 1 | XDR_DROSU  | XANTHINE DEHYDROGENASE | 1.87e+01 |
| 32 | 48 | 66.7 | 655  | 1 | HGFA_HUMAN | HEPATOCYTE GROWTH FACT | 2.83e+01 |
| 33 | 48 | 66.7 | 683  | 1 | APCE_SYNP6 | PHYCOBILISOME LINKER P | 2.83e+01 |
| 34 | 48 | 66.7 | 1131 | 1 | PHYA_SOYBN | PHYTOCHROME A          | 2.83e+01 |
| 35 | 48 | 66.7 | 1353 | 1 | XDR_CALVI  | XANTHINE DEHYDROGENASE | 2.83e+01 |
| 36 | 48 | 66.7 | 1354 | 1 | PUR4_DROME | PHOSPHORIBOSYLFORMYLGL | 2.83e+01 |
| 37 | 47 | 65.3 | 239  | 1 | RL7_DROME  | 60S RIBOSOMAL PROTEIN  | 4.24e+01 |
| 38 | 47 | 65.3 | 485  | 1 | YCGO_ECOLI | HYPOTHETICAL 51.4 KD P | 4.24e+01 |
| 39 | 47 | 65.3 | 764  | 1 | XJ30_YEAST | HYPOTHETICAL 87.2 KD P | 4.24e+01 |
| 40 | 47 | 65.3 | 1112 | 1 | PHYE_ARATH | PHYTOCHROME E          | 4.24e+01 |
| 41 | 47 | 65.3 | 1128 | 1 | PHY3_AVESA | PHYTOCHROME A TYPE 3   | 4.24e+01 |
| 42 | 47 | 65.3 | 1129 | 1 | PHYB_SOLTV | PHYTOCHROME B          | 4.24e+01 |
| 43 | 47 | 65.3 | 1131 | 1 | PHY_PINSY  | PHYTOCHROME            | 4.24e+01 |
| 44 | 47 | 65.3 | 1136 | 1 | PHY_PICAB  | PHYTOCHROME            | 4.24e+01 |
| 45 | 47 | 65.3 | 1156 | 1 | PHYB_SOYBN | PHYTOCHROME B          | 4.24e+01 |

## ALIGNMENTS

| RESULT   | 1  | STANDARD:  | PRT: | 503 AA. |
|--|--|------------|------|---------|
| ID   | Y226_MYCPN   |            |      |         |
| AC   | P75462:  |            |      |         |
| DT   | 01-NOV-1997 (REL. 35, CREATED)                                   |            |      |         |
| DT   | 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)                      |            |      |         |
| DT   | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)                    |            |      |         |
| DE   | HYPOTHETICAL PROTEIN MG226 HOMOLOG.                              |            |      |         |
| OS   | MYCOPLASMA PNEUMONIAE.   |            |      |         |
| OC   | PROKARYOTA: TENERICTES; MOLICUTES; MYCOPLASMA; MYCOPLASMATALES;  |            |      |         |
| OC   | MYCOPLASMATALEAE.  |            |      |         |
| RN   | [1]  |            |      |         |
| RP   | SEQUENCE FROM N.A.   |            |      |         |
| RC   | STRAIN-ATCC 29342 / M129;  |            |      |         |
| RX   | MEDLINE; 97105885.   |            |      |         |
| RA   | HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,      |            |      |         |
| RA   | HERRMANN R.;   |            |      |         |
| RL   | NUCLEIC ACIDS RES. 24:4420-4449(1996).                           |            |      |         |
| CC   | -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL). |            |      |         |
| CC   | -1- SIMILARITY: TO M.GENITALIUM MG225.                           |            |      |         |
| DR   | EMBL; AE000051; G1674212; -                                      |            |      |         |
| KW   | HYPOTHETICAL PROTEIN; TRANSMEMBRANE.                             |            |      |         |
| FT   | TRANSMEM 20 40   | POTENTIAL. |      |         |
| FT   | TRANSMEM 43 63   | POTENTIAL. |      |         |
| FT   | TRANSMEM 106 126   | POTENTIAL. |      |         |
| FT   | TRANSMEM 138 158   | POTENTIAL. |      |         |
| FT   | TRANSMEM 166 186   | POTENTIAL. |      |         |
| FT   | TRANSMEM 215 235   | POTENTIAL. |      |         |
| FT   | TRANSMEM 249 269   | POTENTIAL. |      |         |
| FT   | TRANSMEM 301 321   | POTENTIAL. |      |         |
| FT   | TRANSMEM 359 379   | POTENTIAL. |      |         |
| FT   | TRANSMEM 405 425   | POTENTIAL. |      |         |
| FT   | TRANSMEM 443 463   | POTENTIAL. |      |         |
| FT   | TRANSMEM 468 488   | POTENTIAL. |      |         |
| SQ   | SEQUENCE 503 AA; 54960 MW; 72E63CBC CRC32;                       |            |      |         |
| Query Match  |  |            |      |         |
| Best Local Similarity 55.6%; Score 52; DB 1; Length 503;   |  |            |      |         |
| Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0; |  |            |      |         |
| DB   | 23 FAFAVYAGF 31  |            |      |         |
| Qy   | 1 FPFKYAAAF 9  |            |      |         |
| RESULT   | 2  | STANDARD:  | PRT: | 529 AA. |
| ID   | NMT_AJECA  |            |      |         |
| AC   | P34763:  |            |      |         |

DT 01-FEB-1994 (REL. 28, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
 DE GLYCYPEPTIDE N-TETRADECANOLTRANSFERRASE (EC 2.3.1.97) (PEPTIDE  
 DE N-MYRISTOYLTRANSFERASE) (MYRISTOYL-COA:PROTEIN N-MYRISTOYLTRANSFERASE)  
 DE (NMT)  
 OS ATELLOMYCES CAPSULATA (HISTOPLASMA CAPSULATUM).  
 CC EUKARYOTA; FUNGI; ASCOMYCOTINA; PLECTOMYCETES; GYMNASCALES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-G217B.  
 RX MEDLINE: 94132075.  
 RA LODGE J.K., JOHNSON R.L., WEINBERG R.A., GORDON J.I.;  
 RL J. BIOL. CHEM. 269:2996-3009(1994).  
 CC -1- FUNCTION: ADDS MYRISTOYL GROUP TO N-TERMINAL GLYCINE RESIDUE  
 CC OR CERTAIN CELLULAR AND VIRAL PROTEINS.  
 CC -1- CATALYTIC ACTIVITY: TETRADECANOL-COA + GLYCYL-PEPTIDE - COA +  
 CC N-TETRADECANOYLGLYCYL-PEPTIDE.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: STRONG, TO NMT IN OTHER SPECIES.  
 DR EMBL: L25118; G407695; -.  
 DR PROSITE: PS00975; NMT\_1; 1.  
 DR PROSITE: PS00976; NMT\_2; 1.  
 KM TRANSFERRASE: ACYLTRANSFERASE.  
 SQ SEQUENCE 529 AA; 5363 MW; 6B6ED646 CRC32;  
 Query Match 72.2%; Score 52; DB 1; Length 529;  
 Best Local Similarity 55.6%; Pred. No. 5.26e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Db 192 FRNYSRAF 200  
 Oy 1 FPFYAAAF 9  
 RESULT 3  
 ID ARYA RANGE STANDARD; PRT; 702 AA.  
 AC P14296;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
 DE ARYLPHORIN ALPHA SUBUNIT PRECURSOR.  
 OS MANDUCA SEXTA (TOBACCO HAWMOTH) (TOBACCO HORNWORM).  
 CC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; LEPIDOPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LARVAL FAT BODY;  
 RX MEDLINE: 90037032.  
 RA WILLOTT E., MANG X.-Y., WELLS M.A.;  
 RL J. BIOL. CHEM. 264:19052-19059(1989).  
 CC -1- FUNCTION: ARYLPHORIN IS A LARVAL STORAGE PROTEIN (LSP) WHICH MAY  
 CC SERVE AS A STORAGE PROTEIN USED PRIMARILY AS A SOURCE OF AROMATIC  
 CC AMINO ACIDS FOR PROTEIN SYNTHESIS DURING METAMORPHOSIS. IT IS A  
 CC CONSTITUENT OF THE SCLEROTIZING SYSTEM OF THE CUTICLE, AND SERVES  
 CC AS A CARRIER FOR ECDYSTEROID HORMONE.  
 CC -1- SUBUNIT: ARYLPHORIN IS AN HEXAMER OF SUBUNITS ALPHA AND BETA.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- TISSUE-SPECIFICITY: FAT BODY.  
 CC -1- SIMILARITY: TO ARYA, TO B.MORI STORAGE PROTEINS 1 AND 2, AND TO  
 CC ARTHROPOD HEMOCYANINS.  
 DR EMBL: M28394; G159485; -.  
 DR EMBL: M28396; G159487; -.  
 DR PIR: A34434; A34434.  
 DR HSSP: P04253; ILIA.  
 DR PROSITE: PS00209; HEMOCYANIN\_1; 1.  
 DR PROSITE: PS00210; HEMOCYANIN\_2; 1.  
 KM SIGNAL; STORAGE PROTEIN; GLYCOPROTEIN; MULTIGENE FAMILY.  
 FT SIGNAL 1 16  
 FT CHAIN 17 702 ARYLPHORIN ALPHA SUBUNIT.  
 FT CARBOHYD 75 75 POTENTIAL.  
 FT CARBOHYD 214 214 POTENTIAL.  
 FT SEQUENCE 702 AA; 83866 MW; 5F4B87CD CRC32;  
 SQ

Query Match 72.2%; Score 52; DB 1; Length 702;  
 Best Local Similarity 44.4%; Pred. No. 5.26e+00;  
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Db 686 FPFYENPF 694  
 Oy 1 FPFYAAAF 9  
 RESULT 4  
 ID YBR3 YEAST STANDARD; PRT; 404 AA.  
 AC P38083;  
 DT 01-OCT-1994 (REL. 30, CREATED)  
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 46.4 KD PROTEIN IN ORC2-T1P1 INTERGENIC REGION.  
 GN YBR03C OR YBR0610.  
 OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
 CC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RA DOMDEY H., GASENUEBER H., OBERMAIER B., PIRAVANDI E.;  
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDJ DATA BANKS.  
 DR EMBL: Z35932; G536307; -.  
 DR PIR: S45923; S45923.  
 KM HYPOTHETICAL PROTEIN: TRANSMEMBRANE.  
 FT TRANSMEM 35 55 POTENTIAL.  
 FT TRANSMEM 92 112 POTENTIAL.  
 SQ SEQUENCE 404 AA; 46444 MW; 821F8780 CRC32;  
 Query Match 70.8%; Score 51; DB 1; Length 404;  
 Best Local Similarity 44.4%; Pred. No. 8.09e+00;  
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Db 10 FPFYGSDF 18  
 Oy 1 FPFYAAAF 9  
 RESULT 5  
 ID CSG\_METPE STANDARD; PRT; 593 AA.  
 AC P27373;  
 DT 01-AUG-1992 (REL. 23, CREATED)  
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
 DE CELL SURFACE GLYCOPROTEIN PRECURSOR (S-LAYER PROTEIN).  
 GN SLGA.  
 OS METHANOTHERMUS FERVIDUS.  
 CC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOBACTERIALES; METHANOTHERMACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-42.  
 RC STRAIN-DSM 2088 / V24S;  
 RX MEDLINE: 91293115.  
 RA BROECKL G., BEHR M., FABRY S., HENSEL R., KAUEWITZ H., BIENDL E.,  
 RA KOENIG H.;  
 RL EUR. J. BIOCHEM. 199:147-152(1991).  
 CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY  
 CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIA IS COVERED BY A  
 CC S-LAYER WITH HEXAGONAL SYMMETRY.  
 DR EMBL: X58297; G809714; -.  
 DR PIR: S16225; S16225.  
 DR GLYCOPROTEIN; CELL WALL; S-LAYER; SIGNAL.  
 FT SIGNAL 1 22  
 FT CHAIN 23 593 CELL SURFACE GLYCOPROTEIN.  
 FT CARBOHYD 29 29  
 FT CARBOHYD 58 58 POTENTIAL.  
 FT CARBOHYD 66 66 POTENTIAL.  
 FT CARBOHYD 74 74 POTENTIAL.  
 FT CARBOHYD 114 114 POTENTIAL.  
 FT CARBOHYD 122 122 POTENTIAL.  
 FT CARBOHYD 145 145 POTENTIAL.

FT CARBOHYD 148 148 POTENTIAL.  
 FT CARBOHYD 158 158 POTENTIAL.  
 FT CARBOHYD 176 176 POTENTIAL.  
 FT CARBOHYD 208 208 POTENTIAL.  
 FT CARBOHYD 231 231 POTENTIAL.  
 FT CARBOHYD 326 326 POTENTIAL.  
 FT CARBOHYD 336 336 POTENTIAL.  
 FT CARBOHYD 340 340 POTENTIAL.  
 FT CARBOHYD 431 431 POTENTIAL.  
 FT CARBOHYD 471 471 POTENTIAL.  
 FT CARBOHYD 500 500 POTENTIAL.  
 FT CARBOHYD 516 516 POTENTIAL.  
 SQ SEQUENCE 593 AA: 65481 MW: 5CFA9AA9 CRC32;

Query Match 70.8%; Score 51; DB 1; Length 593;  
 Best Local Similarity 62.5%; Pred. No. 8.09e+00;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 526 YPEKYAVS 533  
 QY 1 FPEKYAAA 8

RESULT 6 STANDARD; PRT; 593 AA.  
 ID CSG\_METSC  
 AC P27374;  
 DT 01-AUG-1992 (REL. 23, CREATED)  
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
 DE CELL SURFACE GLYCOPROTEIN PRECURSOR (S-LAYER PROTEIN).  
 GN SLGA.  
 OS METHANOTHERMUS SOCIABILIS.  
 OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOBACTERIALES; METHANOTHERMACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DSM 3496 / KFI-FL;  
 RX MEDLINE: 91293115.  
 RA BROECKL G., BEHR M., FABRY S., HENSEL R., KAUEWITZ H., BIENDL E.,  
 RA KOENIG H.;  
 RL EUR. J. BIOCHEM. 199;147-152(1991).  
 CC -1- SUBUNIT: ASSEMBLE INTO MONO-LAYERED CRYSTALLINE ARRAYS.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL.  
 DR EMBL: X58296; G809717;  
 DR PIR: S16375; S16375.  
 KW GLYCOPROTEIN; CELL WALL; S-LAYER; SIGNAL.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 593 CELL SURFACE GLYCOPROTEIN.  
 FT CARBOHYD 29 29 POTENTIAL.  
 FT CARBOHYD 58 58 POTENTIAL.  
 FT CARBOHYD 66 66 POTENTIAL.  
 FT CARBOHYD 74 74 POTENTIAL.  
 FT CARBOHYD 114 114 POTENTIAL.  
 FT CARBOHYD 122 122 POTENTIAL.  
 FT CARBOHYD 145 145 POTENTIAL.  
 FT CARBOHYD 148 148 POTENTIAL.  
 FT CARBOHYD 158 158 POTENTIAL.  
 FT CARBOHYD 176 176 POTENTIAL.  
 FT CARBOHYD 208 208 POTENTIAL.  
 FT CARBOHYD 221 221 POTENTIAL.  
 FT CARBOHYD 326 326 POTENTIAL.  
 FT CARBOHYD 336 336 POTENTIAL.  
 FT CARBOHYD 340 340 POTENTIAL.  
 FT CARBOHYD 431 431 POTENTIAL.  
 FT CARBOHYD 471 471 POTENTIAL.  
 FT CARBOHYD 500 500 POTENTIAL.  
 FT CARBOHYD 516 516 POTENTIAL.  
 SQ SEQUENCE 593 AA: 65503 MW: 52B1B8C8 CRC32;

Query Match 70.8%; Score 51; DB 1; Length 593;  
 Best Local Similarity 62.5%; Pred. No. 8.09e+00;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 DB 526 YPEKYAVS 533

QY 1 FPEKYAAA 8

RESULT 7 STANDARD; PRT; 451 AA.  
 ID NMT\_CANAL  
 AC P30418;  
 DT 01-APR-1993 (REL. 25, CREATED)  
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
 DE GLYCYPEPTIDE N-TETRADECANOYLTRANSFERASE (EC 2.3.1.97) (PEPTIDE  
 DE N-MYRISTOYLTRANSFERASE) (MYRISTOYL-COA:PROTEIN N-MYRISTOYLTRANSFERASE)  
 DE (NMT).  
 GN NMT1.  
 OS CANDIDA ALBICANS (YEAST).  
 OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92235090.  
 RA WIEGAND R.C., CARR C., MINNERLY J.C., PAULEY A.M., CARRON C.P.,  
 RA LANGNER C.A., DURONTO R.J., GORDON J.I.;  
 RL J. BIOL. CHEM. 267:8591-8598(1992).  
 CC -1- FUNCTION: ADS MYRISTOYL GROUP TO N-TERMINAL GLYCINE RESIDUE  
 CC OF CERTAIN CELLULAR AND VIRAL PROTEINS.  
 CC -1- CATALYTIC ACTIVITY: TETRADECANOYL-COA + GLYCYL-PEPTIDE = COA +  
 CC N-TETRADECANOYLGLYCYL-PEPTIDE  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: STRONG, TO NMT IN OTHER SPECIES.  
 DR EMBL: M80544; G170884;  
 DR PIR: A38099; A38099.  
 DR PROSITE: PS00975; NMT\_1; 1.  
 DR PROSITE: PS00976; NMT\_2; 1.  
 KW TRANSFERASE; ACYLTRANSFERASE.  
 SQ SEQUENCE 451 AA: 51877 MW: 52BD42D9 CRC32;

Query Match 69.4%; Score 50; DB 1; Length 451;  
 Best Local Similarity 55.6%; Pred. No. 1.24e+01;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 115 FRFKISHEF 123  
 QY 1 FPEKYAAF 9

RESULT 8 STANDARD; PRT; 549 AA.  
 ID YJCG\_ECOLI  
 AC P32705;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 59.2 KD PROTEIN IN SOX-R-ACS INTERGENIC REGION (F549).  
 GN YJCG.  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE: 94089392.  
 RA BLATTNER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J.;  
 RA DANIELS D.L.;  
 RL NUCLEIC ACIDS RES. 21:5408-5417(1993).  
 CC -1- SUBCELLULAR LOCATION: INTERAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE SODIUM: Solute symporter family (SSF).  
 DR EMBL: U00006; G396402;  
 DR EMBL: AE000480; G1790503;  
 DR ECOGENE: EG11942; YJCG.  
 DR PROSITE: PS00456; NA\_SOLUT\_SYM\_1; 1.  
 DR PROSITE: PS00457; NA\_SOLUT\_SYM\_2; 1.  
 KW HYPOTHETICAL PROTEIN; TRANSPORT; TRANSMEMBRANE; SODIUM TRANSPORT;  
 SYMPORT.  
 FT TRANSMEM 33 24 POTENTIAL.  
 FT TRANSMEM 33 53 POTENTIAL.

|    |     |          |     |
|----|-----|----------|-----|
| Db | 494 | EPYKPALE | 502 |
|    |     | 11: 1: 1 |     |
| QY | 1   | EPFKYAAF | 9   |

```

D01 111 LPYFAAF 119
      :4::111
OY 1 FPEKYAAF 9

```

```

DE 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN MG366 HOMOLOG.
OS MYCOPLASMA PNEUMONIAE.
CC PROKARYOTA: TENERICUTES; MOLICUTES; MYCOPLASMA; MYCOPLASMALES;
OC MYCOPLASMAACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE: 97105885.
RA HIMMELREICH R., HILBERT H., FLAGENS H., PIRKL E., LI B.-C.,
RA HERMANN R.;
RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
DR EMBL: AE000027; G1673969; -.
SQ HYPOTHETICAL PROTEIN.
KW SEQUENCE 664 AA; 76769 MW; E4D94A05 CRC32;

Query Match 69.4%; SCORE 50; DB 1; Length 664;
Best Local Similarity 44.4%; Pred. No. 1,24e+01;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 358 FAYKSEIF 366
Qy 1 FPKYIAAF 9

RESULT 11
ID VPH1 YEAST STANDARD; PRT; 840 AA.
AC P32563;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE VACUOLAR ATP SYNTHASE 95.5 KD SUBUNIT (EC 3.6.1.34).
GN VPH1 OR YOR270C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92332542.
RA MANOLSON M.F., PROTEAU D., PRESTON R.A., STENBIT A., ROBERTS B.T.,
RA HOTI M.A., PREUS D., MULHOLLAND J., BOTSSTEIN D., JONES E.W.;
RL J. BIOL. CHEM. 267:14294-14303(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93147685.
RA MANOLSON M.F., PROTEAU D., JONES E.W.;
RL J. EXP. BIOL. 172:105-112(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C.
RX MEDLINE: 97051594.
RA CHERET G., BERNARDI A., SOR F.J.;
RL YEAST 12:1059-1064(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE: 97298311.
RA POIREY R., TAUNTAUX J.C.;
RL YEAST 13:483-487(1997).
CC -1- FUNCTION: RODUTED FOR ASSEMBLY AND ACTIVITY OF THE VACUOLAR
APPAE. POTENTIAL ROLE IN DIFFERENTIAL TARGETING AND REGULATION OF
THE ENZYME FOR A SPECIFIC ORGANELLE.
CC -1- SIMILARITY: BELONGS TO THE VAPFASE 116 KD SUBUNIT FAMILY.
CC -1- SIMILARITY: BELONGS TO THE VAPFASE 116 KD SUBUNIT FAMILY.
DR EMBL: M89778; G173173; -.
DR EMBL: X89633; E189397; -.
DR EMBL: Z75178; E252118; -.
PIR: A42970; A42970.
SGD: L0002467; VPH1.
KW HYDROGEN ION TRANSPORT; TRANSMEMBRANE; GLYCOPROTEIN.
RN [5]
DT DOMAIN 1 406
RT TRANSMEM 407 441
RT 1 (POTENTIAL).

```



```

Db      461 PPMWFAIME 469
      |||::| |
QY      1 PPKRYAAAF 9

RESULT 13
      WNT5_1PROME STANDARD; PRT; 1010 AA.
AC      P28466; Q01535;
DT      01-DEC-1992 (REL. 24, CREATED)
DT      01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT      01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE      PROTEIN DWNT-5 PRECURSOR (DWNT-3).
GN      WNT-5 OR WNT-3.
OS      DROSOPHILA MELANOGASTER ('FRUIT FLY').
OC      EUKARYOTA; METAEOA; ARTHROPODA; INSECTA; DIPTERA
      [1]
RP      SEQUENCE FROM N.A.
RP      STRAIN=CANTON-S;
RC      MEDLINE; 93048811.
RA      RUSSELL J.; GENNISSEN A.; NUSSE R.;
RL      DEVELOPMENT 115:475-482(1992).
      [2]
RP      SEQUENCE FROM N.A.
RP

```

| ID | NAME   | STANDARD | PRT        | 1010 AA              |
|----|--|----------|------------|----------------------|
| AC | P28466; 001535;  |          |            |                      |
| DT | 01-DEC-1992 (REL. 24, CREATED)                                   |          |            |                      |
| DT | 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)                      |          |            |                      |
| DT | 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)                    |          |            |                      |
| DE | PROTEIN DWNT-5 PRECURSOR (DWNT-3).                               |          |            |                      |
| GN | WNT-5 OR WNT-3.  |          |            |                      |
| OS | DROSOPHILA MELANOGASTER (FRUIT FLY).                             |          |            |                      |
| OC | EUDARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.                |          |            |                      |
| RN | [1]  |          |            |                      |
| RP | SEQUENCE FROM N.A.   |          |            |                      |
| RC | STRAIN=CANTON-S.   |          |            |                      |
| RX | MEDLINE; 93048811.   |          |            |                      |
| RA | RUSSELL J., GENNISSEN A., NUSSE R.;                              |          |            |                      |
| RL | DEVELOPMENT 115:475-482(1992).                                   |          |            |                      |
| RZ | [2]  |          |            |                      |
| RP | SEQUENCE FROM N.A.   |          |            |                      |
| RX | MEDLINE; 93050786.   |          |            |                      |
| RA | EISENBERG L.M., INGHAM P.W., BROWN A.M.C.;                       |          |            |                      |
| RL | DEV. BIOL. 154:73-83(1992).                                      |          |            |                      |
| CC | -1- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING |          |            |                      |
| CC | MOLECULE WHICH AFFECTS THE DEVELOPMENT OF DISCRETE REGIONS OF    |          |            |                      |
| CC | TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS. MAY   |          |            |                      |
| CC | HAVE A ROLE IN THE CNS DEVELOPMENT.                              |          |            |                      |
| CC | -1- SUBCELLULAR LOCATION: SECRETED (PROBABLE).                   |          |            |                      |
| CC | -1- SIMILARITY: TO OTHER MEMBERS OF THE WNT FAMILY.              |          |            |                      |
| DR | EMBL; X64736; G7907; -.  |          |            |                      |
| DR | EMBL; M97450; G158806; -.  |          |            |                      |
| DR | FLYBASE; FBgn0010194; Wnt5.                                      |          |            |                      |
| DR | PROSITE; PS00246; WNT1.1.  |          |            |                      |
| KW | DEVELOPMENTAL PROTEIN; GLYCOPROTEIN; SIGNAL.                     |          |            |                      |
| FT | SIGNAL   | 1        | 28         | POTENTIAL.           |
| FT | CHAIN  | 29       | 1010       | PROTEIN DWNT-5.      |
| FT | DOMAIN   | 280      | 288        | POLY-GLY.            |
| FT | DOMAIN   | 461      | 476        | POLY-SER.            |
| FT | CARBOHYD   | 60       | 60         | POTENTIAL.           |
| FT | CARBOHYD   | 66       | 66         | POTENTIAL.           |
| FT | CARBOHYD   | 115      | 115        | POTENTIAL.           |
| FT | CARBOHYD   | 219      | 219        | POTENTIAL.           |
| FT | CARBOHYD   | 310      | 310        | POTENTIAL.           |
| FT | CARBOHYD   | 344      | 344        | POTENTIAL.           |
| FT | CARBOHYD   | 425      | 425        | POTENTIAL.           |
| FT | CARBOHYD   | 490      | 490        | POTENTIAL.           |
| FT | CARBOHYD   | 491      | 491        | POTENTIAL.           |
| FT | CARBOHYD   | 534      | 534        | POTENTIAL.           |
| FT | CARBOHYD   | 599      | 599        | POTENTIAL.           |
| FT | CARBOHYD   | 730      | 730        | POTENTIAL.           |
| FT | CARBOHYD   | 958      | 958        | POTENTIAL.           |
| FT | CONFLICT   | 281      | 283        | MISSING (IN REF. 2). |
| FT | CONFLICT   | 320      | 320        | E -> D (IN REF. 2).  |
| FT | CONFLICT   | 474      | 476        | MISSING (IN REF. 2). |
| SO | SEQUENCE   | 1010 AA; | 112875 MW; | 507BD98C CRC32;      |

Query Match 69.4%; Score 50; DB 1; Length 1010;

Best Local Similarity 44.4%; Pred. No. 1.24e+01;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 673 FAYKATDF 681

1 FPKYAAAF 9

| RESULT | 14  | STANDARD; | PRT; | 1379 AA. |
|--------|---|-----------|------|----------|
| ID     | MET MOUSE                                   |           |      |          |
| AC     | P16056:                                     |           |      |          |
| DT     | 01-APR-1990 (REL. 14, CREATED)              |           |      |          |
| DT     | 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE) |           |      |          |

01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE HEPATOCYTE GROWTH FACTOR RECEPTOR PRECURSOR (MET PROTO-ONCOGENE  
 DE TYROSINE KINASE) (EC 2.7.1.112) (HGF-SF RECEPTOR).  
 GN MET.  
 OS MUS MUSCULUS (MOUSE).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 88262253.  
 RA CHAN A.M.L., KING H.W.S., DEAKIN E.A., TEMPEST P.R., HILKENS J.,  
 RA KROMBERG V., EDWARDS D.R., WILLS A.J., BROOKES P., COOPER C.S.,  
 RA ONCOGENE 2:593-599(1988).  
 CC -1- FUNCTION: RECEPTOR FOR HEPATOCYTE GROWTH FACTOR. HAS A TYROSINE-  
 CC PROTEIN KINASE ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -1- SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN (50 KD) AND A BETA  
 CC CHAIN (145 KD) WHICH ARE DISULFIDE LINKED.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- DISEASE: ACTIVATION OF MET AFTER REARRANGEMENT WITH THE TPR  
 CC (TRANSLOCATED PROMOTER) LOCUS OF CHROMOSOME 1 PRODUCES AN  
 CC ONCOGENIC PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF RECEPTOR.  
 CC EMBL: Y00671; G53059; .  
 DR PIR: S01254; S01254.  
 DR MGD: MGI:96969; MET.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW TRANSFERASE; TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; ATP-BINDING;  
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 1379 HEPATOCYTE GROWTH FACTOR RECEPTOR.  
 FT DOMAIN 25 931 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 932 954 POTENTIAL.  
 FT DOMAIN 955 1379 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 1076 1343 PROTEIN KINASE.  
 FT SITE 306 307 CLEAVAGE (POTENTIAL).  
 FT NE-BIND 1082 1090. ATP (BY SIMILARITY).  
 FT BINDING 1108 1108. ATP (BY SIMILARITY).  
 FT ACT\_SITE 1202 1202. BY SIMILARITY.  
 FT MOD\_RES 1233 1233. PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 45 45. POTENTIAL.  
 FT CARBOHYD 106 106. POTENTIAL.  
 FT CARBOHYD 201 201. POTENTIAL.  
 FT CARBOHYD 357 357. POTENTIAL.  
 FT CARBOHYD 398 398. POTENTIAL.  
 FT CARBOHYD 404 404. POTENTIAL.  
 FT CARBOHYD 606 606. POTENTIAL.  
 FT CARBOHYD 634 634. POTENTIAL.  
 FT CARBOHYD 784 784. POTENTIAL.  
 FT CARBOHYD 878 878. POTENTIAL.  
 SQ SEQUENCE 1379 AA; 153548 MW; E1597F1A CRC32;

Query Match 69.4%; Score 50; DB 1; Length 1379;  
 Best Local Similarity 55.6%; Pred. No. 1.24e+01;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 364 FPIKYVNDP 372  
 11:11  
 1 FPKRYAANF 9

RESULT 15  
 ID MET\_HUMAN STANDARD; PRT; 1390 AA.  
 AC P08581;  
 DT 01-AUG-1988 (REL. 08, CREATED)  
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HEPATOCYTE GROWTH FACTOR RECEPTOR PRECURSOR (MET PROTO-ONCOGENE  
 DE TYROSINE KINASE) (EC 2.7.1.112) (HGF-SF RECEPTOR).  
 GN MET.

OS HOMO SAPIENS (HUMAN).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA GIORDANO S.;  
 RL SUBMITTED (NOV-1990) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87317655.  
 RA PARK M., DEAN M., KAUL K., BRAUN M.J., GONDA M.A., VANDE WOUDE G.;  
 RA PROC. NATL. ACADEM. SCI. U.S.A. 84:6379-6383(1987).  
 RN [3]  
 RP SEQUENCE OF 1010-1390 FROM N.A.  
 RX MEDLINE: 88143699.  
 RA CHAN A.M.L., KING H.W.S., TEMPEST P.R., DEAKIN E.A., COOPER C.S.,  
 RA BROOKES P.;  
 RL ONCOGENE 1:229-233(1987).  
 RN [4]  
 RP SEQUENCE OF 1267-1390 FROM N.A.  
 RX MEDLINE: 86065462.  
 RA DEAN M., PARK M., LE BEAU M.M., ROBINS T.S., DIAZ M.O., ROWLEY J.D.,  
 RA BLAIR D.G., VANDE WOUDE G.F.;  
 RL NATURE 318:385-388(1985).  
 RN [5]  
 RP FUNCTION.  
 RX MEDLINE: 91118019.  
 RA BOTTARO D.P., RUBIN J.S., FALETTO D.L., CHAN A.M.-L., KMECIK T.E.,  
 RA VANDE WOUDE G.F., ARONSON S.A.;  
 RL SCIENCE 251:802-804(1991).  
 RN [6]  
 RP PHOSPHORYLATION AT TYR-1235.  
 RX MEDLINE: 92011756.  
 RA FERRACINI R., LONGATTI P., NALDINI L., VIGNA E., COMOGILIO P.M.;  
 RA J. BIOL. CHEM. 266:19558-19564(1991).  
 RN [7]  
 RP VARIANTS HPRC, AND VARIANT VAL-320.  
 RX MEDLINE: 9285124.  
 RA SCHMIDT L., DUH F.-M., CHEN F., KISHIDA T., GLENN G., CHOYE P.,  
 RA SCHERER S.W., ZHUANG Z., LOBENSKI I., DEAN M., ALLIKMETS R.,  
 RA CHIDMABARAW A., BERGERHEIM U.R., FELTIS J.T., CASADEVAL C.,  
 RA ZAMARRON A., BERNES M., RICHARD S., LIPS C.J.M., WALTHER M.M.,  
 RA TSUI L.-C., GELL L., ORCUTT M.L., STACKHOUSE T., LIPAN J., STIFE L.,  
 RA BRAUCH H., DECKER J., NIEHANS G., HUGHSON M.D., MOCH H., STORKEL S.,  
 RA LERMAN M.I., LINEHAN W.M., ZBAR B.;  
 RL NAT. GENET. 16:68-73(1997).  
 CC -1- FUNCTION: RECEPTOR FOR HEPATOCYTE GROWTH FACTOR. HAS A TYROSINE-  
 CC PROTEIN KINASE ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -1- SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN (50 KD) AND A BETA  
 CC CHAIN (145 KD) WHICH ARE DISULFIDE LINKED.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- DISEASE: ACTIVATION OF MET AFTER REARRANGEMENT WITH THE TPR  
 CC GENE PRODUCES AN ONCOGENIC PROTEIN.  
 CC -1- DISEASE: DEFECTS IN MET ARE THE CAUSE OF HEREDITARY PAPILLARY  
 CC RENAL CARCINOMA (HPRC). HPRC IS A FORM OF INHERITED KIDNEY CANCER  
 CC CHARACTERIZED BY A PREDISPOSITION TO DEVELOP MULTIPLE, BILATERAL  
 CC PAPILLARY RENAL TUMOURS. THE PATTERN OF INHERITANCE IS CONSISTENT  
 CC WITH AUTOSOMAL DOMINANT TRANSMISSION WITH REDUCED PENETRANCE.  
 CC -1- SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF RECEPTOR.  
 CC EMBL: M35074; G386868; .  
 DR EMBL: X54559; .; NOT\_ANNOTATED\_CDS.  
 DR EMBL: J02958; G307196; .  
 DR PIR: A40175; TVHWU.  
 DR MIM: 164860; .  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW TRANSFERASE; TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; ATP-BINDING;  
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION; SIGNAL;  
 KW CHROMOSOMAL TRANSLOCATION; DISEASE MUTATION; POLYMORPHISM.  
 FT SIGNAL 1 24

FT CHAIN 25 1390 HEPATOCTE GROWTH FACTOR RECEPTOR.  
FT DOMAIN 25 932 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 933 955 POTENTIAL.  
FT DOMAIN 956 1390 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 1078 1345 PROTEIN KINASE.  
FT NP\_BIND 1084 1092 ATP (BY SIMILARITY).  
FT BINDING 1110 1110 ATP (BY SIMILARITY).  
FT ACT\_SITE 1204 1204 BY SIMILARITY.  
FT SITE 307 308 CLEAVAGE (POTENTIAL).  
FT SITE 1009 1010 BREAKPOINT FOR TRANSLLOCATION TO FORM  
TFR-MET ONCOGENE.  
FT MOD\_RES 1235 1235 PHOSPHORYLATION (AUTO-).  
FT CARBOHYD 45 45 POTENTIAL.  
FT CARBOHYD 106 106 POTENTIAL.  
FT CARBOHYD 149 149 POTENTIAL.  
FT CARBOHYD 202 202 POTENTIAL.  
FT CARBOHYD 399 399 POTENTIAL.  
FT CARBOHYD 405 405 POTENTIAL.  
FT CARBOHYD 607 607 POTENTIAL.  
FT CARBOHYD 635 635 POTENTIAL.  
FT CARBOHYD 785 785 POTENTIAL.  
FT CARBOHYD 879 879 POTENTIAL.  
FT CARBOHYD 930 930 POTENTIAL.  
FT VARIANT 320 320 A -> V.  
FT VARIANT 1131 1131 M -> T (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1188 1188 V -> L (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1195 1195 L -> V (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1220 1220 V -> I (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1228 1228 D -> N (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1230 1228 D -> H (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1230 1230 Y -> C (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1230 1230 Y -> H (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1250 1250 M -> T (IN HPRC; SOMATIC MUTATION).  
FT CONFLICT 755 755 S -> STMWKEPLNIVSFLCERAS (IN REF. 2).  
FT CONFLICT 1191 1191 G -> A (IN REF. 2).  
SQ SEQUENCE 1390 AA: 155526 MW: 650992C2 CRC32:

## Query Match

Best Local Similarity 55.68%; Score 50; DB 1; Length 1390;  
Pred. No. 1.24e+01;

Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 365 FPIKYVND 373

1 FPFKYAAF 9

Search completed: Fri Sep 11 12:38:51 1998  
Job time : 8 secs.

**THIS PAGE BLANK (USPTO)**

# MPSEARCH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch.p protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 11 12:39:09 1998; Maspar time 4.15 Seconds

Tabular output not generated. 91.424 Million cell updates/sec

Title: >US-08-452-843-1  
Description: (1-9) from US08452843.pep  
Perfect Score: 72  
Sequence: 1 FPFKAAAF 9

Scoring table:  
PAM 150  
Gap 15

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: 1:sp.fungi 2:sp.human 3:sp.invertebrate 4:sp.mammal  
5:sp.mhc 6:sp.organelle 7:sp.phage 8:sp.plant  
9:sp.bacteria 10:sp.todent 11:sp.virus 12:sp.vertibrate  
13:sp.unclassified

Statistics: Mean 23.284; Variance 34.517; scale 0.675

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                   | Pred. No. |
|------------|-------|-------------|--------|-------|-------------------------------|-----------|
| 1          | 55    | 76.4        | 173    | 9     | O26436 CONSERVED PROTEIN.     | 3.59e+00  |
| 2          | 55    | 76.4        | 469    | 9     | P73738 HYPOTHETICAL 52.5 KD P | 3.59e+00  |
| 3          | 55    | 76.4        | 800    | 3     | O21145 KOZELL.1               | 5.59e+00  |
| 4          | 54    | 75.0        | 466    | 3     | O16976 TO2B11.6 PROTEIN.      | 5.44e+00  |
| 5          | 54    | 75.0        | 975    | 9     | O46147 ALPHA-TOXIN (FRAGMENT) | 5.44e+00  |
| 6          | 54    | 75.0        | 2178   | 9     | O46149 ALPHA-TOXIN.           | 5.44e+00  |
| 7          | 54    | 75.0        | 4578   | 12    | O42181 PKD1 PROTEIN.          | 5.44e+00  |
| 8          | 52    | 72.2        | 828    | 3     | O23370 ZC518.3.               | 1.23e+01  |
| 9          | 51    | 70.8        | 139    | 3     | O23322 ZC443.2.               | 1.84e+01  |
| 10         | 51    | 70.8        | 211    | 9     | O26053 HYPOTHETICAL 24.9 KD P | 1.84e+01  |
| 11         | 50    | 69.4        | 88     | 9     | O05234 HYPOTHETICAL 10.1 KD P | 2.73e+01  |
| 12         | 50    | 69.4        | 117    | 2     | O13863 DNA-BINDING PROTEIN.   | 2.73e+01  |
| 13         | 50    | 69.4        | 133    | 10    | O63964 P190-C-MET (FRAGMENT)  | 2.73e+01  |
| 14         | 50    | 69.4        | 321    | 3     | O18046 T06C12.11.             | 2.73e+01  |
| 15         | 50    | 69.4        | 440    | 3     | O17358 COSEA.6 PROTEIN.       | 2.73e+01  |
| 16         | 50    | 69.4        | 490    | 9     | O25863 NADH-UBIQUINONE OXIDOR | 2.73e+01  |
| 17         | 50    | 69.4        | 511    | 9     | O06460 BETA SUBUNIT OF NITRAT | 2.73e+01  |
| 18         | 50    | 69.4        | 1382   | 10    | P97579 HEPATOCYTE GROWTH FACT | 2.73e+01  |
| 19         | 50    | 69.4        | 1382   | 10    | P97573 HGF RECEPTOR PRECURSOR | 2.73e+01  |
| 20         | 49    | 68.1        | 72     | 11    | O84669 GENOME, PARTIAL SEQUEN | 4.04e+01  |

|    |    |      |      |    |                                |          |
|----|----|------|------|----|--------------------------------|----------|
| 21 | 49 | 68.1 | 159  | 8  | O23750 POLLEN ALLERGEN, BETV1  | 4.04e+01 |
| 22 | 49 | 68.1 | 159  | 8  | O23746 POLLEN ALLERGEN, BETV1  | 4.04e+01 |
| 23 | 49 | 68.1 | 160  | 8  | O96366 POLLEN ALLERGEN BET V   | 4.04e+01 |
| 24 | 49 | 68.1 | 160  | 8  | O42499 MAJOR ALLERGEN BET V    | 4.04e+01 |
| 25 | 49 | 68.1 | 160  | 8  | O96370 POLLEN ALLERGEN BET V   | 4.04e+01 |
| 26 | 49 | 68.1 | 160  | 8  | O96365 POLLEN ALLERGEN BET V   | 4.04e+01 |
| 27 | 49 | 68.1 | 160  | 8  | O96367 POLLEN ALLERGEN BET V   | 4.04e+01 |
| 28 | 49 | 68.1 | 160  | 8  | O24642 POLLEN ALLERGEN BETV1   | 4.04e+01 |
| 29 | 49 | 68.1 | 160  | 8  | O39431 MAJOR ALLERGEN BET V    | 4.04e+01 |
| 30 | 49 | 68.1 | 160  | 8  | O23752 POLLEN ALLERGEN BETV1   | 4.04e+01 |
| 31 | 49 | 68.1 | 160  | 8  | O23753 POLLEN ALLERGEN BETV1   | 4.04e+01 |
| 32 | 49 | 68.1 | 160  | 8  | O23751 POLLEN ALLERGEN BETV1   | 4.04e+01 |
| 33 | 49 | 68.1 | 160  | 8  | O96371 POLLEN ALLERGEN BET V   | 4.04e+01 |
| 34 | 49 | 68.1 | 209  | 3  | O01769 COSMID 2C581.           | 4.04e+01 |
| 35 | 49 | 68.1 | 291  | 9  | O28106 4-HYDROXYBENZONATE OCTA | 4.04e+01 |
| 36 | 49 | 68.1 | 232  | 12 | O92038 ACYL-COA DESATURASE 1   | 4.04e+01 |
| 37 | 49 | 68.1 | 391  | 9  | O49633 HYPOTHETICAL 42.1 KD P  | 4.04e+01 |
| 38 | 49 | 68.1 | 545  | 11 | O90054 ORE2.                   | 4.04e+01 |
| 39 | 49 | 68.1 | 607  | 9  | O30286 KIMS.                   | 4.04e+01 |
| 40 | 49 | 68.1 | 822  | 9  | O30286 MOLYBDOPROTEIN OXIDORE  | 4.04e+01 |
| 41 | 49 | 68.1 | 2470 | 12 | O90681 CATION-DEPENDENT MAN    | 4.04e+01 |
| 42 | 48 | 66.7 | 585  | 8  | O24044 LACCASE (EC 1.10.3.2).  | 5.94e+01 |
| 43 | 48 | 66.7 | 586  | 8  | O24043 LACCASE (EC 1.10.3.2).  | 5.94e+01 |
| 44 | 48 | 66.7 | 640  | 3  | O02015 PAX-A.                  | 5.94e+01 |
| 45 | 48 | 66.7 | 1139 | 9  | O54073 ANCHOR PROTEIN, LCM.    | 5.94e+01 |

## ALIGNMENTS

| RESULT   | ID  | 1                                       | PRELIMINARY:                          | PRT: | 173 AA. |
|--|---|---|---------------------------------------|------|---------|
| AC   | O26436  |   |                                       |      |         |
| DT   | 01-JAN-1998   | (TREMBLREL. 05, CREATED)                |                                       |      |         |
| DT   | 01-JAN-1998   | (TREMBLREL. 05, LAST SEQUENCE UPDATE)   |                                       |      |         |
| DE   | 01-JAN-1998   | (TREMBLREL. 05, LAST ANNOTATION UPDATE) |                                       |      |         |
| DE   | CONSERVED PROTEIN.  |   |                                       |      |         |
| GN   | MT336.  |   |                                       |      |         |
| OS   | METHANOBACTERIUM THERMAUTOTROPHICUM.                                    |   |                                       |      |         |
| OC   | ARCHAEBACTERIA, EURYARCHAEOTA, METHANOBACTERIALES;                      |   |                                       |      |         |
| OC   | METHANOBACTERIACEAE.  |   |                                       |      |         |
| RN   | [1]   |   |                                       |      |         |
| RP   | SEQUENCE FROM N.A.  |   |                                       |      |         |
| RC   | STRAIN-DELTA H.   |   |                                       |      |         |
| RA   | SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,   |   |                                       |      |         |
| RA   | ALDRIDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,           |   |                                       |      |         |
| RA   | HARRISON D., HOANG L., KEAGLE P., LUMM W., POTLIER B., QIU D.,          |   |                                       |      |         |
| RA   | SPARAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R., JIWANI N., |   |                                       |      |         |
| RA   | CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S., MCDONAGALL S.,  |   |                                       |      |         |
| RA   | SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M., DANIELS C.J.,          |   |                                       |      |         |
| RA   | MAO J.-I., RICE P., NOLLING J., REBEVE J.N.,                            |   |                                       |      |         |
| RL   | J. BACTERIOL. 179:7135-7155(1997).                                      |   |                                       |      |         |
| RN   | [2]   |   |                                       |      |         |
| RP   | SEQUENCE FROM N.A.  |   |                                       |      |         |
| RC   | STRAIN-DELTA H.   |   |                                       |      |         |
| RA   | SMITH D.R.:   |   |                                       |      |         |
| RL   | SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.                    |   |                                       |      |         |
| DR   | EMBL: A5000818; G2621392; -   |   |                                       |      |         |
| DR   | SEQUENCE 173 AA; 20426 MW; 131A50A2 CRC32;                              |   |                                       |      |         |
| Query Match  |   |   |                                       |      |         |
| Best local similarity 76.4%; Score 55; DB 9; Length 173;   |   |   |                                       |      |         |
| Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0; |   |   |                                       |      |         |
| DB   | 32 FPFKAAAF 40  |   |                                       |      |         |
| OY   | 1 FPFKAAAF 9  |   |                                       |      |         |
| RESULT   | 2   |   |                                       |      |         |
| ID   | P73738  |   | PRELIMINARY:                          | PRT: | 469 AA. |
| AC   | P73738;   |   |                                       |      |         |
| DT   | 01-FEB-1997   |   | (TREMBLREL. 02, CREATED)              |      |         |
| DT   | 01-FEB-1997   |   | (TREMBLREL. 02, LAST SEQUENCE UPDATE) |      |         |

DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 52.5 KD PROTEIN.  
 GN NDH.  
 OS SYNECHOCYSTIS SP.  
 OC EUBACTERIA: CYANOBACTERIA: CHROCOCCALES: SYNECHOCYSTIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RA TABATA S.;  
 RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RA YANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARIO K.,  
 RA OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A.,  
 RA YAMADA M., YASUDA M., TABATA S.;  
 RL DNA RES. 3:109-116(1996).  
 DR EMBL: D90909; G1652868; .  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 469 AA; 52544 MW; 0887A65E CRC32;  
 Query Match 76.4%; Score 55; DB 9; Length 469;  
 Best Local Similarity 55.6%; Pred. No. 3.59e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Db 282 FPFKPSKF 290  
 OY 1 FPFKTAAP 9  
 RESULT 3 PRELIMINARY; PRT: 800 AA.  
 ID 021145;  
 AC 021145;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE KOZEIL.1.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA: METAZOA: ACCELLOMATES: NEMATODA: SECERNENTEA: RHABDITIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MCMURRAY A.;  
 RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE; 94150718.  
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,  
 RA GARDNER A., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GRADNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
 RL NATURE 368:32-38(1994).  
 DR EMBL: 277665; E256003; .  
 SQ SEQUENCE 800 AA; 90383 MW; A52285D2 CRC32;  
 Query Match 76.4%; Score 55; DB 3; Length 800;  
 Best Local Similarity 75.0%; Pred. No. 3.59e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Db 529 FPFYTNNA 536  
 OY 1 FPFYTAAP 8  
 RESULT 4 PRELIMINARY; PRT: 466 AA.  
 ID 016976;

AC 016976;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE TO2B11.6 PROTEIN.  
 GN TO2B11.6.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA: METAZOA: ACCELLOMATES: NEMATODA: SECERNENTEA: RHABDITIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE; 94150718.  
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,  
 RA GRADNER A., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
 RL NATURE 368:32-38(1994).  
 DR EMBL: AF022979; G2384879; .  
 SQ SEQUENCE 466 AA; 51764 MW; 557E8291 CRC32;  
 Query Match 75.0%; Score 54; DB 3; Length 466;  
 Best Local Similarity 66.7%; Pred. No. 5.44e+00;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 Db 302 LPEFGAAP 310  
 OY 1 FPFYTAAP 9  
 RESULT 5 PRELIMINARY; PRT: 975 AA.  
 ID 046147;  
 AC 046147;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE ALPHA-TOXIN (FRAGMENT).  
 OS CLOSTRIDIUM NOVI.  
 OC PROKARYOTA: BACTERIA: FIRMICUTES: ENDOSPORE-FORMING RODS AND COCCI;  
 OC BACILLACEAE: CLOSTRIDIUM.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC19402;  
 RA HOFMANN F., HABERMANN E., VON EICHEL-STREIBER C.;  
 RL SUBMITTED (MAY-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: 223280; G479121; .  
 FT NON\_TER 1  
 SQ SEQUENCE 975 AA; 112285 MW; B1FF18A7 CRC32;  
 Query Match 75.0%; Score 54; DB 9; Length 975;  
 Best Local Similarity 66.7%; Pred. No. 5.44e+00;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 70 FPMKEAPF 78  
 OY 1 FPFYTAAP 9

```

RESULT 6
ID 046149 PRELIMINARY: PRT: 2178 AA.
AC 046149;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DE 2518.3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RA THOMAS K.;
RA WILKINSON-SPROAT J.; WOHLDMAN P.;
RA VAUGHAN K.; WATERSTON R.; WATSON A.; WEINSTOCK L.;
RA WILKINSON-SPROAT J.; WOHLDMAN P.;
RA NATURE 368:32-38(1994).
DR EMBL: 268753; E218668;
SQ SEQUENCE 2178 AA; 250134 MW; 0C347C36 CRC32;

Query Match
Best Local Similarity 66.7%; Pred. No. 5.44e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1273 FFWKTEAPF 1281
OY 1 FPFKYAAAF 9

RESULT 7
ID 042181 PRELIMINARY: PRT: 4578 AA.
AC 042181;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DE 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE PKD1 PROTEIN.
GN PKD1.
OS FUGU RUBRIPES (JAPANESE PUFFERFISH) (TAKIFUGU RUBRIPES).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA.
OC OSTEICHTHYES; ACTINOPTERYGII; TETRAODONTIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 97449170.
RA SANDFORD R., SCOTTO B., APARICIO S., BRENNER S., VAUDIN M., WILSON R.,
RA CHRISOE S., PEPIN K., BATEMAN A., CHOTHA C., HUGHES J., HARRIS P.;
RA HUM. MOL. GENET. 6:1483-1489(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA VAUDIN M.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WASHU;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF013614; G2627436;
SQ SEQUENCE 4578 AA; 504585 MW; DE8EE954 CRC32;

Query Match
Best Local Similarity 75.0%; Score 54; DB 9; Length 4578;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 2951 FPFNTVANY 2959
OY 1 FPFKYAAAF 9

RESULT 8
ID 023370 PRELIMINARY: PRT: 828 AA.
AC 023370;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

```

```

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE 2518.3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RA THOMAS K.;
RA WILKINSON-SPROAT J.; WOHLDMAN P.;
RA VAUGHAN K.; WATERSTON R.; WATSON A.; WEINSTOCK L.;
RA WILKINSON-SPROAT J.; WOHLDMAN P.;
RA NATURE 368:32-38(1994).
DR EMBL: 268753; E218668;
SQ SEQUENCE 828 AA; 94021 MW; 266AAD6D CRC32;

Query Match
Best Local Similarity 72.2%; Score 52; DB 3; Length 828;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 434 FPFNTVANY 442
OY 1 FPFKYAAAF 9

RESULT 9
ID 023322 PRELIMINARY: PRT: 139 AA.
AC 023322;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE 26443.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RA BAYNES C.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J.,
RA COULSON A., CRAYTON M., DEAR S., DU Z., DUBBIN R., FAVELLO A.,
RA FULTON L., GARDNER A., GREEN P., HAKINS T., HILTER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
RA LAURELLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIEKEN L., ROOPRA A.,
RA SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONNHAMMER E.,
RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J.; WOHLDMAN P.;
RA NATURE 368:32-38(1994).
DR EMBL: 275553; E250312;
SQ SEQUENCE 139 AA; 16819 MW; 1C24D160 CRC32;

Query Match
Best Local Similarity 70.8%; Score 51; DB 3; Length 139;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 66 FPFNTVANY 74
OY 1 FPFKYAAAF 9

```

OY 1 PFKYAAAF 9

RESULT 10  
ID 026053 PRELIMINARY; PRT: 211 AA.  
AC 026053:

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 24.9 KD PROTEIN.  
GN HP1525.

OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA.  
RC AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=26695;  
RA TOMB, WHITE, KERLAJAGE, CLAYTON, SUTTON, FLEISCHMANN, KETCHUM, KLENK, GILL, DOUGHERTY, NELSON, QUACKENBUSH, ZHOU, KIRKNESS, PETERSON, LOFTUS, RICHARDSON, DOBSON, KHALAK, GLODEK, MCKENNEY, FITZGERALD, LEE, ADAMS, HICKEY, BERG, GOCAYNE, UTTERBACK, PETERSON, KELLEY, COTTON, WEIDMAN, FUJII, BOWMAN, WATTHEY, WALLIN, HAYES, BORODOVSKY, KARP, SMITH, FRASER VENTER.  
RA NATURE 388:539-547(1997).  
RL [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=26695;  
RA TOMB, WHITE, KERLAJAGE, CLAYTON, SUTTON, FLEISCHMANN, KETCHUM, KLENK, GILL, DOUGHERTY, NELSON, QUACKENBUSH, ZHOU, KIRKNESS, PETERSON, LOFTUS, RICHARDSON, DOBSON, KHALAK, GLODEK, MCKENNEY, FITZGERALD, LEE, ADAMS, HICKEY, BERG, GOCAYNE, UTTERBACK, PETERSON, KELLEY, COTTON, WEIDMAN, FUJII, BOWMAN, WATTHEY, WALLIN, HAYES, BORODOVSKY, KARP, SMITH, FRASER VENTER.  
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AE000650; G2314707; -  
KM HYPOTHETICAL PROTEIN.  
SO SEQUENCE 211 AA; 24866 MW; 6BED7882 CRC32;

Query Match  
Best Local Similarity 70.4%; Score 51; DB 9; Length 211;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 189 FAFYDSAF 197  
OY 1 PFKYAAAF 9

RESULT 11  
ID 005234 PRELIMINARY; PRT: 88 AA.  
AC 005234:

DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 10.1 KD PROTEIN.  
GN YUGE.  
OS BACILLUS SUBTILIS.  
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA OUDERA B., KONINGSSTEYN G.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA DANCHIN A.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA KUNST F., OGASAMARA N., MOSER I., ALBERTINI A.M., ALLONI G., AZEVEDO V., BERTEHO M.G., BESSIERES P., SOLOTTIN A., BORCHERT S., BORRIS R., BOURSTIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,

RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M., CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A., DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T., RA ENTIAN K.D., ERRINGTON J., FABBET C., FERRARI E., FOUGER D., RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N., RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G., RA GUISSEPI G., GUY B.J., HAGA K., HALECH J., HARWOOD C.R., HENAUT A., RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L., RA JORIS B., KARAWATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C., RA KOBAASHI Y., KOETTER P., KONINGSSTEIN G., KROGH S., KIMANO M., RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAAREVIC V., RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C., RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M., RA NOONE D., O'REILLY M., OGAMA K., OGIMARA A., OUDEGA B., PARK S.H., RA PARO V., POHL T.M., PORTELETTE D., PORMOLIK S., PRESCOTT A.M., RA PRESERVAN E., PUTIC P., PURNELLE B., RAPPOPORT G., REY M., REYNOLDS S., RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SDALE Y., RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOPFONE F., RA SERIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDO B., RA SOROKIN A., TACCONE E., TAKAGI T., TAKAHASHI H., TAKEKAWA K., RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI D., RA TOSANO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A., RA VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZEGGER T., RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUDOTO K., YATA K., RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;  
RL NATURE 390:249-256(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA KUNST F., OGASAMARA N., YOSHIKAWA H., DANCHIN A.;  
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Z93934; E311524; -  
KM EMBL: Z93120; E1184221; -  
SO HYPOTHETICAL PROTEIN.  
SO SEQUENCE 88 AA; 10065 MW; FD93F758 CRC32;

Query Match  
Best Local Similarity 69.4%; Score 50; DB 9; Length 88;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 21 PFRYGEF 28  
OY 2 PFKYAAAF 9

RESULT 12  
ID 013863 PRELIMINARY; PRT: 117 AA.  
AC 013863:

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE DNA-BINDING PROTEIN.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 95194400.  
RA LUZI P., STRAYER D.S.;  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 208:153-160(1995).  
DR EMBL: L10405; G860730; -  
KM DNA-BINDING.  
SO SEQUENCE 117 AA; 12849 MW; 1CF727F5 CRC32;

Query Match  
Best Local Similarity 69.4%; Score 50; DB 2; Length 117;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

DB 52 FSKYSAT 59  
OY 1 PFKYAAAF 8



```

RESULT 13
ID 063984 PRELIMINARY; PRT: 132 AA.
AC 063984
GN
DE 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE P190-C-MET (FRAGMENT).
GN C-MET
OS RATIUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94220137.
RA TSUJII M., KAWANO S., TSUJI S., ITO T., HAYASHI N., HORIMOTO M.,
RA MITA E., NAGANO K., MASUDA E., HAYASHI N.;
RA BIOCHEM. BIOPHYS. RES. COMMUN. 200:536-541(1994).
DR EMBL: S69881; G546428.
FT NON TER
SQ SEQUENCE 132 AA; 15027 MW; 79F6F69F CRC32;

Query Match
Best Local Similarity 55.6%; Pred. No. 2.73e+01;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 3 FPKIYVDF 11
11:11
OY 1 FPKYAAF 9

RESULT 14
ID 018046 PRELIMINARY; PRT: 321 AA.
AC 018046
GN
DE 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE T06C12.11.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACLOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RA KELLY P.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., LAISTER N., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULLSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(0).
DR EMBL: 281116; F1198242.
SQ SEQUENCE 321 AA; 36630 MW; 89E0CE24 CRC32;

Query Match
Best Local Similarity 50.0%; Pred. No. 2.73e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 77 PFRYPGF 84
11:11
OY 2 PFRYAAF 9

RESULT 15
ID 017358 PRELIMINARY; PRT: 440 AA.
AC 017358
GN
DE 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

```

```

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE C05E4.6 PROTEIN.
GN C05E4.6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACLOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., LAISTER N., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULLSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2.
RA BLANCHARD M., KRAMER J., GIBSON A.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2.
RA WATSON R.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF026209; G2435568.
SQ SEQUENCE 440 AA; 50630 MW; 50E5F923 CRC32;

Query Match
Best Local Similarity 50.0%; Pred. No. 2.73e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 77 PFRYPGF 84
11:11
OY 2 PFRYAAF 9

```

Search completed: Fri Sep 11 12:39:49 1998  
Job time : 40 secs.

**THIS PAGE BLANK (USPTO)**